

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:34:09 ; Search time 22.3488 Seconds
(without alignments)
1735.006 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 403
Sequence: 1 MLHHVKLIIVATSKRLVGKK.....MKKRELARIWDEIEKKLS 403

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 283416 seqs, 96216763 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	72	17.9	401	G71018	pantothenate metab
2	60	14.9	401	E75114	dna/pantothenate m
3	12	3.0	394	A72223	pantothenate metab
4	11	2.7	403	A64414	pantothenate metab
5	10	2.5	394	D81819	hypothetical prote
6	10	2.5	394	A81058	DNA/pantothenate m
7	10	2.5	399	A11302	pantothenate metab
8	10	2.5	399	A11674	pantothenate metab
9	9	2.2	404	D69455	pantothenate metab
10	8	2.0	127	T23443	hypothetical prote
11	8	2.0	210	S62355	high mobility grou
12	8	2.0	265	A99249	hypothetical prote
13	8	2.0	298	B97226	dihydroorotate deh
14	8	2.0	391	E71125	probable aromatic-
15	8	2.0	396	T36548	probable transamin
16	8	2.0	434	AF1238	E. coli YbdN prote
17	8	2.0	434	AB1601	E. coli YbdN prote
18	8	2.0	480	1 TVMTG	phosphoglycerate k
19	8	2.0	481	D71927	cag island protein
20	8	2.0	481	B64585	cag pathogenicity
21	8	2.0	794	T36972	probable membrane
22	8	2.0	1261	S75130	sensory transducti
23	7	1.7	36	T03348	gene e19 protein -
24	7	1.7	60	G69115	hypothetical prote
25	7	1.7	97	E97156	exonuclease VII sm
26	7	1.7	91	F69252	hypothetical prote
27	7	1.7	101	F90925	hypothetical prote
28	7	1.7	101	B85774	hypothetical prote
29	7	1.7	111	F70338	methionine-trNA li

30	7	1.7	124	2	C70537	hypothetical prote
31	7	1.7	130	2	F75075	hypothetical prote
32	7	1.7	138	2	C64382	hypothetical prote
33	7	1.7	140	2	E75056	probable translati
34	7	1.7	140	2	D71104	probable translati
35	7	1.7	144	2	A96580	hypothetical prote
36	7	1.7	145	2	A97722	50S ribosomal prot
37	7	1.7	145	2	E71723	ribosomal protein
38	7	1.7	149	2	D97333	hypothetical prote
39	7	1.7	160	2	T10278	protein tyrosine p
40	7	1.7	170	2	T43779	ribosomal protein
41	7	1.7	173	2	G82818	disulfide bond for
42	7	1.7	183	2	H95142	flavoprotein limpo
43	7	1.7	183	2	F98010	N-terminal region
44	7	1.7	186	2	B61611	nonhistone chromos
45	7	1.7	190	2	C90260	hypothetical prote

ALIGNMENTS

RESULT 1

G71018
pantothenate metabolism flavoprotein dfp homolog PH1444 - Pyrococcus horikoshii
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C:Accession: G71018
R:Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguch
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: G71018
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-401 <KAW>
A:Cross-references: UNIPROT:O59114; GB:AP000006; NID:g33236133; PIDN:BAA30551.1; PID:g32
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1444

Query Match 17.9%; Score 72; DB 2; Length 401;

Best Local Similarity 100.0%; Pred. No. 6.6e-63;

Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 TGNPVITETGTGFIHVELAGEHENKADLLVCPATANTISKIACGIDDPVTVTVTAPP 127

DB 69 TGNPVITETGTGFIHVELAGEHENKADLLVCPATANTISKIACGIDDPVTVTVTAPP 128

QY 128 HIPIMIAPAMHE 139

DB 129 HIPIMIAPAMHE 140

RESULT 2

E75114

dna/pantothenate metabolism flavoprotein (dfp) PAB1897 - Pyrococcus abyssi (strain Orsa
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: E75114

R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str

A:Reference number: A75001

A:Accession: E75114

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-401 <KAW>

A:Cross-references: UNIPROT:Q9V0S3; GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB496;

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1897
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 14.9%; Score 60; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 5e-51;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 ISEVELAGSHENKADLILVCPATANTISKIAGIDDPVTVTTFAPHIPIMIAPMHE 139
|||||
DB 81 ISEVELAGSHENKADLILVCPATANTISKIAGIDDPVTVTTFAPHIPIMIAPMHE 140
|||||

RESULT 3
A72223
pantothenate metabolism flavoprotein dfp homolog TM1687 - Thermotoga maritima (strain MS)
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: A72223
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.

Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: A72223
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <ARN>
A:Cross-references: UNIPROT:O9X213; GB:AE001809; GB:AE000512; NID:g4982257; PIDN:AAD3675
A:Experimental source: strain MS98
C:Genetics:
A:Gene: TM1687
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 3.0%; Score 12; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 RPIITNASSGKMG 223
|||||
DB 200 RPIITNASSGKMG 211
|||||

RESULT 4
A64414
pantothenate metabolism flavoprotein MJ0913 - Methanococcus jannaschii
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: A64414
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: A64414
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-403 <BUL>
A:Cross-references: UNIPROT:Q58323; GB:U67535; GB:L77117; NID:g2826348; PIDN:AAB98918.1;
C:Genetics:
A:Map position: REV845792-844581
C:Superfamily: pantothenate metabolism flavoprotein dfp
C:Keywords: flavoprotein

Query Match 2.7%; Score 11; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSGKMGVALAE 228
|||||

DB 219 SSGKMGVALAE 229
|||||

RESULT 5
DB1819
hypothetical protein NMA1916 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
C:Accession: DB1819
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel;
; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: AB1775; MUID:20222556; PMID:10761919
A:Accession: DB1819
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <PAR>
A:Cross-references: UNIPROT:Q9J798; GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB8511
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA1916
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 2.5%; Score 10; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSGKMGVALA 227
|||||
DB 208 SSGKMGVALA 217
|||||

RESULT 6
AB1058
DNA/pantothenate metabolism flavoprotein NMB1658 [imported] - Neisseria meningitidis (st
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: AB1058
R:Tettelin, H.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
Hickey, E.K.; Hsu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755; PMID:10710307
A:Accession: AB1058
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <TET>
A:Cross-references: UNIPROT:Q9JVB7; GB:AE002516; GB:AE002098; NID:g7326905; PIDN:AAF4201
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1658
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 2.5%; Score 10; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSGKMGVALA 227
|||||
DB 208 SSGKMGVALA 217
|||||

RESULT 7
AB1302
pantothenate metabolism flavoprotein homolog lmo1825 [imported] - Listeria mono
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AB1302
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker;

```

.; Dominguez-Bernal, G.; Duchoaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: A11302
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-399 <GLA>
A:Cross-references: UNIPROT:Q8Y674; GB:NC_003210; PIDN:CAC99903.1; PID:g16411279; GSPDB:
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo1825
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 2.5%; Score 10; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVTAGATRE 206
|||||
Db 190 VLVTAGATRE 199

RESULT 8
A11674
pantothenate metabolism flavoprotein homolog lin1939 [imported] - Listeria innoc
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: A11674
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchoaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: A11674
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-399 <GLA>
A:Cross-references: UNIPROT:Q92AI3; GB:AL592022; PIDN:CAC97169.1; PID:g16414440; GSPDB:G
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin1939
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 2.5%; Score 10; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVTAGATRE 206
|||||
Db 190 VLVTAGATRE 199

RESULT 9
D69455
pantothenate metabolism flavoprotein dfp homolog AF1645 - Archaeoglobus fulgidus
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: D69455
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Wosse, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea

```

```

A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: D69455
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-404 <MLE>
A:Cross-references: UNIPROT:Q28628; GB:AE000989; GB:AE000782; MID:g269312; PIDN:AAB895
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 2.2%; Score 9; DB 2; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 TANTISKIA 110
|||||
Db 103 TANTISKIA 111

RESULT 10
T23443
hypothetical protein K08C9.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T23443
R:Lennard, N.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19741
A:Accession: T23443
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-127 <WIL>
A:Cross-references: UNIPROT:Q9XU72; EMBL:Z81567; PIDN:CAB04589.1; GSPDB:GN00019; CESP:K
A:Experimental source: clone K08C9
C:Genetics:
A:Gene: CESP:K08C9.6
A:Map position: 1
A:Introns: 52/2, 69/1
C:Superfamily: Caenorhabditis elegans hypothetical protein K08C9.6

Query Match 2.0%; Score 8; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 FRIRIKL 255
|||||
Db 85 FRIRIKL 92

RESULT 11
S62355
high mobility group protein 1 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: S62355
R:Nightingale, K.; Dimitrov, S.; Reeves, R.; Wolffe, A.P.
EMBO J. 15, 548-561, 1996
A:Title: Evidence for a shared structural role for HMGI and linker histones B4 and H1
A:Reference number: S62355; MUID:96174815; PMID:8599938
A:Accession: S62355
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-210 <NIG>
A:Cross-references: UNIPROT:Q91596; EMBL:U21933; MID:g709959; PIDN:AAC59859.1; PID:g709
A>Note: the authors did not translate the codon for residue 1
C:Superfamily: nonhistone chromosomal protein HMGI-2; HMGI box homology
F:6-92/Domain: HMGI box homology <HMGI>
F:91-165/Domain: HMGI box homology <HMGI>

Query Match 2.0%; Score 8; DB 2; Length 210;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 SDFRPKIK 293
|||||

```

Db 106 SDFRPKIK 113

RESULT 12

A99249

hypothetical protein fabG-1 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C:Accession: A99249

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweyez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.
arrett, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: A99249

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-265 <KUR>

A:Cross-references: UNIPROT:Q97ZE4; GB:AE006641; NID:G13814158; PIDN:AAK41248.1; GSPDB:G

C:Genetics:

A:Gene: fabG-1

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 2.0%; Score 8; DB 2; Length 265;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 GKRVLVTA 201
|||||
Db 7 GKRVLVTA 14

RESULT 13

B97226

dihydroorotate dehydrogenase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004

C:Accession: B97226

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97226

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-298 <KUR>

A:Cross-references: UNIPROT:Q97FS7; GB:AE001437; PIDN:AAK80597.1; PID:G15025678; GSPDB:G

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2650

C:Superfamily: dihydroorotate oxidase

Query Match 2.0%; Score 8; DB 2; Length 298;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 KELPKMKK 387
|||||
Db 82 KELPKMKK 89

RESULT 14

E71125

probable aromatic-amino-acid transaminase (EC 2.6.1.57) PH0771 [similarity] - Pyrococcus

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004

C:Accession: E71125

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A:Reference number: A71000; MUID:98344137; PMID:9679194

A:Accession: E71125

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-391 <KAW>

A:Cross-references: UNIPROT:O58489; GB:AP000003; NID:G3236130; PIDN:BAA29863.1; PID:G325

A:Experimental source: strain OT3

A>Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH0771

C:Keywords: aminotransferase; phosphoprotein; pyridoxal phosphate

F:236/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 2.0%; Score 8; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 TSKEKLIIE 342
|||||
Db 371 TSKEKLIIE 378

RESULT 15

T36548

probable transaminase (EC 2.6.1.-) SCH10.36 [similarity] - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: T36548

R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999

A:Reference number: Z21609

A:Accession: T36548

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-396 <MUR>

A:Cross-references: UNIPROT:Q9X8S5; EMBL:AL049754; PIDN:CAB42045.1; GSPDB:GN00070; SCOE

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOE:DB: SCH10.36

C:Superfamily: aspartate transaminase

C:Keywords: aminotransferase

Query Match 2.0%; Score 8; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVVTAGAT 204
|||||
Db 99 VLVVTAGAT 106

Search completed: July 2, 2005, 12:38:48
Job time : 24.3488 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:22:48 ; Search time 10.6047 Seconds
(without alignments)
1098.127 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPDWKIRKEILIEPFSEE.....PYRGNVQGSTRLAFSKRKL 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A-COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B-COMB.pep.*
3: /cgn2_6/prodata/1/iaa/6A-COMB.pep.*
4: /cgn2_6/prodata/1/iaa/6B-COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCUTS-COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	806	100.0	156	4	US-09-399-003-71
2	533	66.1	246	3	US-08-822-774-44
3	533	66.1	246	3	US-09-632-711-44
4	533	66.1	246	3	US-09-632-703B-44
5	533	66.1	246	4	US-09-632-702-44
6	533	66.1	246	4	US-09-399-003-44
7	185.5	23.0	197	4	US-09-328-352-4665
8	172	21.3	202	4	US-09-252-991A-17662
9	168.5	20.9	192	4	US-09-540-236-3681
10	168.5	20.9	204	4	US-09-543-681A-5063
11	164	20.3	217	4	US-09-489-039A-8299
12	160	19.9	42	3	US-08-822-774-41
13	160	19.9	42	3	US-09-632-711-41
14	160	19.9	42	3	US-09-632-703B-41
15	160	19.9	42	4	US-09-632-702-41
16	160	19.9	42	4	US-09-399-003-41
17	138	17.1	190	4	US-09-198-452A-412
18	138	17.1	190	4	US-09-438-185A-394
19	134.5	16.7	177	3	US-09-199-637A-247
20	121	15.0	147	4	US-09-583-110-4437
21	121	15.0	158	4	US-09-107-433-3527
22	119	14.8	27	3	US-08-822-774-38
23	119	14.8	27	3	US-09-632-711-38
24	119	14.8	27	3	US-09-632-703B-38
25	119	14.8	27	4	US-09-632-702-38
26	119	14.8	27	4	US-09-399-003-38
27	118	14.6	151	4	US-09-902-540-10618

28	114.5	14.2	138	2	US-08-824-405-11	Sequence 11, Appl
29	112	13.9	181	3	US-09-199-637A-239	Sequence 239, App
30	111	13.8	138	2	US-08-824-405-10	Sequence 10, Appl
31	111	13.8	164	2	US-08-824-405-2	Sequence 2, Appl
32	111	13.8	252	2	US-08-824-405-4	Sequence 4, Appl
33	111	13.8	252	4	US-09-538-092-1062	Sequence 1062, Ap
34	111	13.8	280	4	US-09-949-016-10934	Sequence 10934, A
35	102	12.7	178	3	US-09-171-461-2	Sequence 2, Appl
36	102	12.7	178	4	US-09-970-711-2	Sequence 2, Appl
37	99.5	12.3	254	3	US-09-199-637A-241	Sequence 241, App
38	87.5	10.9	183	3	US-09-134-001C-5336	Sequence 5336, Ap
39	87.5	10.9	1150	3	US-09-238-303-9	Sequence 9, Appl
40	87.5	10.9	1150	4	US-09-946-239-9	Sequence 9, Appl
41	87	10.8	147	2	US-08-824-405-12	Sequence 12, Appl
42	87	10.8	177	4	US-09-248-796A-17657	Sequence 17657, A
43	87	10.8	277	4	US-08-311-731A-284	Sequence 284, App
44	85.5	10.6	1031	3	US-08-811-682-15	Sequence 15, Appl
45	84	10.4	475	3	US-09-134-001C-4250	Sequence 4250, Ap

ALIGNMENTS

RESULT 1
US-09-399-003-71
; Sequence 71, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Hogrefe, Holly
; APPLICANT: Hansen, Connie J
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
; TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying
; FILE REFERENCE: 4121.0116-02
; CURRENT APPLICATION NUMBER: US/09/399,003
; CURRENT FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/ US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-399-003-71

Query Match 100.0%; Score 806; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 7.8e-93;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFAVKGLIDVEKGVVIPPYVALI	60
Db	1	MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFAVKGLIDVEKGVVIPPYVALI	60
QY	61	LTLEIRKLPDDVGMGDKIRSSLAEGVIGSFVWDPCWGNLTMLYNASNEPELVRYGE	120
Db	61	LTLEIRKLPDDVGMGDKIRSSLAEGVIGSFVWDPCWGNLTMLYNASNEPELVRYGE	120
QY	121	RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL	156
Db	121	RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL	156

RESULT 2
US-08-822-774-44
; Sequence 44, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)

; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-632-703B-44

Query Match 66.1%; Score 533; DB 3; Length 246;
Best Local Similarity 98.1%; Pred. No. 2.5e-58;
Matches 102; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 53 PPREVALIILTLERIKLPDDVMGDMKIRSSSLAREGVIGSAWVDPGWDGNLTMLYNASNE 112
Db 1 PPREVALIILTLERIKLPNNVGMGDMKIRSSSLAREGVIGSAWVDPGWDGNLTMLYNASNE 60

QY 113 PVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156
Db 61 PVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 104

RESULT 5
US-09-632-702-44
; Sequence 44, Application US/09632702
; Patent No. 6444428
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,702
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-632-702-44

Query Match 66.1%; Score 533; DB 4; Length 246;
Best Local Similarity 98.1%; Pred. No. 2.5e-58;
Matches 102; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 53 PPREVALIILTLERIKLPDDVMGDMKIRSSSLAREGVIGSAWVDPGWDGNLTMLYNASNE 112
Db 1 PPREVALIILTLERIKLPNNVGMGDMKIRSSSLAREGVIGSAWVDPGWDGNLTMLYNASNE 60

QY 113 PVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156
Db 61 PVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 104

RESULT 6
US-09-399-003-44
; Sequence 44, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Hansen, Holly
; APPLICANT: Hogrefe, Connie J
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
; and Methods for Purifying and Identifying
; TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying
; FILE REFERENCE: 4121.0116-02
; CURRENT APPLICATION NUMBER: US/09/399,003
; CURRENT FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/ US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(246)
; OTHER INFORMATION: "X" represents any amino acid
US-09-399-003-44

Query Match 66.1%; Score 533; DB 4; Length 246;
Best Local Similarity 98.1%; Pred. No. 2.5e-58;
Matches 102; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 53 PPREVALIILTLERIKLPDDVMGDMKIRSSSLAREGVIGSAWVDPGWDGNLTMLYNASNE 112
Db 1 PPREVALIILTLERIKLPNNVGMGDMKIRSSSLAREGVIGSAWVDPGWDGNLTMLYNASNE 60

QY 113 PVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156
Db 61 PVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 104

RESULT 7
US-09-328-352-4665
; Sequence 4665, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4665
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4665

Query Match 23.0%; Score 185.5; DB 4; Length 197;
Best Local Similarity 27.4%; Pred. No. 6.4e-15;
Matches 48; Conservative 31; Mismatches 63; Indels 33; Gaps 4;

QY 10 KEILTEPPFSEESIQ-----PAGYDLRVGRE-----AFVKGK 40
Db 23 KHGMIEPYAENQVRFDXNGEKLIISVSSYGYDVCARBEKVFVTNVHSAIVDPKTFDEKS 82


```
Db 35 WLDEGLAIPRPVRRINGATVDVRLGNKFTFRGHTAPFDLSPKAEVSALDRVMS 94
Qy 47 EGKVIIP-----PREVALITLRIKLPDDVMGDMKIRSLAREGVIGSFA--WVDP 96
Db 95 E-EIVLPEGEAFPLHGCALAVTYESVTLPADLVGWLDCRSSLARLGLMVHTAHRI DP 153
Qy 97 GWDGNLTMLYNASNEPVRLRGFRVQIAFIRLEGPARNPY 138
Db 154 GWSGCVLFEFYNKGLPLALRPGMPIGALSFEPLSGPAARPY 195

RESULT 12
US-08-822-774-41
; Sequence 41, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-822-774-41
Query Match 19.9%; Score 160; DB 3; Length 42;
Best Local Similarity 96.9%; Pred. No. 1e-12;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MLLPDWKIRKEILIEPFSESLQAGYDLRVG 32
Db 11 MLLPDWKIRKEILIEPFSEWLQAGYDLRVG 42

RESULT 13
US-08-632-711-41
; Sequence 41, Application US/09632711
; Patent No. 6333165
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: <Unknown>
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/09/632,703B
; FILING DATE: 24-Aug-2000
```

```
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/09/632,711
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/822,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-632-711-41
Query Match 19.9%; Score 160; DB 3; Length 42;
Best Local Similarity 96.9%; Pred. No. 1e-12;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MLLPDWKIRKEILIEPFSESLQAGYDLRVG 32
Db 11 MLLPDWKIRKEILIEPFSEWLQAGYDLRVG 42

RESULT 14
US-09-632-703B-41
; Sequence 41, Application US/09632703B
; Patent No. 6379553
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF
; Protein Complexes, Isolated PEF Proteins, and Methods for
; Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: <Unknown>
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/09/632,703B
; FILING DATE: 24-Aug-2000
```

```
;
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BARKER, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 4121.0116-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 849-6613
; TELEFAX: (650) 849-6666
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
;
; US-09-632-702-41
;
; Query Match 19.9%; Score 160; DB 3; Length 42;
; Best Local Similarity 96.9%; Pred. No. 1e-12;
; Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVG 32
; | | | | | | | | | | | | | | | | | | | |
; Db 11 MLLPDWKIRKEILIEPFSEEWLQAGYDLRVG 42
;
; Search completed: July 2, 2005, 12:26:00
; Job time : 11.6047 secs
;
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
;
; US-09-632-702-41
;
; Query Match 19.9%; Score 160; DB 4; Length 42;
; Best Local Similarity 96.9%; Pred. No. 1e-12;
; Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVG 32
; | | | | | | | | | | | | | | | | | | | |
; Db 11 MLLPDWKIRKEILIEPFSEEWLQAGYDLRVG 42
;
; Search completed: July 2, 2005, 12:26:00
; Job time : 11.6047 secs
;
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BARKER, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 4121.0116-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 849-6613
; TELEFAX: (650) 849-6666
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
;
; US-09-632-703B-41
;
; Query Match 19.9%; Score 160; DB 3; Length 42;
; Best Local Similarity 96.9%; Pred. No. 1e-12;
; Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVG 32
; | | | | | | | | | | | | | | | | | | | |
; Db 11 MLLPDWKIRKEILIEPFSEEWLQAGYDLRVG 42
;
; RESULT 15
; US-09-632-702-41
; Sequence 41, Application US/09632702
; Patent No. 6444428
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
;
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,702
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:32:14 ; Search time 30.9767 Seconds
(without alignments)
2578.849 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 156
Sequence: 1 MLLPDWKIRKEILIEPFSEE.....PYRGNVQGSTRLAFSKRKL 156

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Uniprot_03.*

1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	156	100.0	156	1 DCD_PYRFU	Q8X251 pyrococcus
2	156	100.0	156	2 Q6T180	Q6T180 pyrococcus
3	37	23.7	156	1 DCD_PYRHO	O57706 pyrococcus
4	22	14.1	154	1 DCD_PYRAB	Q9UXS8 pyrococcus
5	8	5.1	114	2 Q7V7Z6	Q7V7Z6 bordetella
6	8	5.1	114	2 Q7W5U6	Q7W5U6 bordetella
7	8	5.1	114	2 Q7W5U7	Q7W5U7 bordetella
8	8	5.1	420	1 ODO2_BUCAI	P57389 buchnera ap
9	8	5.1	459	2 Q26001	Q26001 plasmidium
10	8	5.1	459	2 Q81501	Q81501 agrobacteri
11	8	5.1	503	2 Q8UBB2	Q8UBB2 agrobacteri
12	7	4.5	34	2 Q9UQN9	Q9UQN9 homo sapien
13	7	4.5	37	2 Q9XQQ5	Q9XQQ5 toxoplasma
14	7	4.5	61	2 Q7P2B1	Q7P2B1 fusobacteri
15	7	4.5	64	1 SCX4_MESMA	P58328 mesobuthus
16	7	4.5	66	2 Q61ZE1	Q61ZE1 mesobuthus
17	7	4.5	84	1 SC13_MESMA	Q9N1C8 mesobuthus
18	7	4.5	84	1 SCXB_MESMA	Q9N1C8 mesobuthus
19	7	4.5	94	2 Q8XTK6	Q8XTK6 raistonla s
20	7	4.5	127	2 Q8DUZ8	Q8DUZ8 streptococc
21	7	4.5	133	1 FLIS_BACSU	Q39739 bacillus su
22	7	4.5	133	2 Q8ENI3	Q8ENI3 oceanobacil
23	7	4.5	133	2 Q9K6W1	Q9K6W1 bacillus ha
24	7	4.5	137	2 Q6PRY2	Q6PRY2 uncultured
25	7	4.5	144	2 Q9HFB0	Q9HFB0 rhizobium l
26	7	4.5	153	2 Q7W8B4	Q7W8B4 bordetella
27	7	4.5	153	2 Q7W1X2	Q7W1X2 bordetella
28	7	4.5	153	2 Q84CS3	Q84CS3 bordetella
29	7	4.5	160	2 Q8MRQ3	Q8MRQ3 drosophila
30	7	4.5	163	2 Q6CU30	Q6CU30 kluyveromyc
31	7	4.5	164	2 Q8CR50	Q8CR50 staphylococ

32 7 4.5 175 2 Q6BIP7 O6BIP7 debaryomyce
33 7 4.5 175 2 Q9UHT7 Q9UHT7 homo sapien
34 7 4.5 177 2 Q8TH89 Q8TH89 methanosarc
35 7 4.5 178 1 AIL_YEREN P16454 yersinia en
36 7 4.5 178 2 Q70AM3 Q70AM3 yersinia en
37 7 4.5 180 2 Q6L1B1 Q6L1B1 picophilus
38 7 4.5 182 1 AIL_YERPS Q56957 yersinia ps
39 7 4.5 182 2 Q667X3 Q667X3 yersinia ps
40 7 4.5 182 2 Q74SP4 Q74SP4 yersinia pe
41 7 4.5 182 2 Q8ZCR3 Q8ZCR3 yersinia pe
42 7 4.5 192 2 Q8EH22 Q8EH22 shewanella
43 7 4.5 194 2 Q8D0Z7 Q8D0Z7 yersinia pe
44 7 4.5 198 2 Q9FMX7 Q9FMX7 arabidopsis
45 7 4.5 198 2 Q7NZY4 Q7NZY4 chromobacte

ALIGNMENTS

RESULT 1

DCD_PYRFU STANDARD; PRT; 156 AA.
ID DCD_PYRFU AC Q8X251;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP deaminase).
GN Name=dcd; OrderedLocuNames=PF1996;
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21664347; PubMed-11782527; DOI=10.1073/pnas.012372799;
RA Hogrefe H.H., Hansen C.J., Scott B.R., Nielson K.B.;
RT "Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation."
RL Proc. Natl. Acad. Sci. U.S.A. 99:596-601(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate family.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AY066005; AAL47572.1; -;
CC EMBL; AE010292; AAL82120.1; -;
CC HSSP; Q57872; LOGH.
CC -----
CC DR HAMAP; MF_00146; -; 1.
CC DR InterPro; IPR003232; dCTP deaminase.
CC DR InterPro; IPR008180; DeoxyUTPase.
CC DR Pfam; PF00692; dUTPase; 1.
CC DR ProDom; PD004900; dCTP deaminase; 1.
CC Complete proteome; Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 156 AA; 17869 MW; 00D57A5A0207B451 CRC64;

Query Match 100.0%; Score 156; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.4e-154;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MLLPDWKIRKEILIBPFSESLQAGYDLRVGREAFVKGLIDVEKEGKVIPPPEYALI 60
DB 1 MLLPDWKIRKEILIBPFSESLQAGYDLRVGREAFVKGLIDVEKEGKVIPPPEYALI 60
QY 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSPAWDPDGDGNLTMLYNASNEPVELRYGE 120
DB 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSPAWDPDGDGNLTMLYNASNEPVELRYGE 120
QY 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSRKKL 156
DB 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSRKKL 156

RESULT 2
ID Q6T180 PRELIMINARY; PRT; 156 AA.
AC Q6T180;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE DUTPase.
OS Pyrococcus woesei.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2262;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22844507; PubMed=12963343; DOI=10.1016/S1046-5928(03)00108-6;
RA Dabrowski S., Klier Ahning B.;
RT "Cloning, expression, and purification of the His6-tagged hyper-
RT thermostable dUTPase from Pyrococcus woesei in Escherichia coli:
RT application in PCR".
RL Protein Expr. Purif. 31:72-78 (2003).
DR EMBL; AY443493; AARI5897.1; -.
DR GO; GO:0046080; P:dUTP metabolism; IEA.
DR InterPro; IPR003232; dCTP deaminase.
DR InterPro; IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
SQ SEQUENCE 156 AA; 17869 MW; 00D57A5A0207B451 CRC64;

Query Match 100.0%; Score 156; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.4e-154;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLPDWKIRKEILIBPFSESLQAGYDLRVGREAFVKGLIDVEKEGKVIPPPEYALI 60
DB 1 MLLPDWKIRKEILIBPFSESLQAGYDLRVGREAFVKGLIDVEKEGKVIPPPEYALI 60
QY 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSPAWDPDGDGNLTMLYNASNEPVELRYGE 120
DB 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSPAWDPDGDGNLTMLYNASNEPVELRYGE 120
QY 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSRKKL 156
DB 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSRKKL 156

RESULT 3
ID DCD_PVRHO STANDARD; PRT; 156 AA.
AC Q57706;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP
DE deaminase).
GN Name=dcd; OrderedLocusNames=PHI1997;
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Maechi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76 (1998).
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000007; BAA31124.1; -.
DR PIR; E71216; E71216.
DR HSSP; Q57872; LOGH.
DR HANAP; MF_00146; -.
DR InterPro; IPR003232; dCTP deaminase.
DR InterPro; IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
KW Complete proteome; Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 156 AA; 17871 MW; 96B2C2C50393A985 CRC64;

Query Match 23.7%; Score 37; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 4.3e-30;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLPDWKIRKEILIBPFSESLQAGYDLRVGREAFV 37
DB 1 MLLPDWKIRKEILIBPFSESLQAGYDLRVGREAFV 37

RESULT 4
ID DCD_PVRAB STANDARD; PRT; 154 AA.
AC Q90XS8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP
DE deaminase).
GN Name=dcd; OrderedLocusNames=PYRAB17800; ORFNames=PAB1164;
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RX MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic V., Porterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512 (2003).
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AJ248288; CAB50685.1; -;
 DR FIR; G75030; G75030.
 DR HSP; Q57872; 10GH.
 DR HAMAP; MF 00146; -; 1.
 DR InterPro; IPR003232; dCTP deaminase.
 DR InterPro; IPR008180; DeoxyUTPase.
 DR Pfam; PF00692; dUTPase; 1.
 DR ProDom; PD004900; dCTP deaminase; 1.
 DR ProComplete proteome; Hydrolyase; Nucleotide metabolism.
 SK SEQUENCE 154 AA; 17758 MW; C031BEE419094DDB CRC64;

Query Match 14.1%; Score 22; DB 1; Length 154;
 Best Local Similarity 100.0%; Pred. No. 2.1e-14; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 GSPAWDPGWDGNLTMLYNAS 110
 |||||
 DB 89 GSPAWDPGWDGNLTMLYNAS 110

RESULT 5
 Q7VYZ6 PRELIMINARY; PRT; 114 AA.
 AC Q7VYZ6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Putative exported protein.
 GN OrderedLocusNames=Bp1153;
 OS Bordetella pertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=520;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640414; CAE41450.1; -;
 DR InterPro; IPR008972; Cupredoxin.
 KW Complete proteome.
 SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;

Query Match 5.1%; Score 8; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 108 NASNEPVE 115
 |||||
 DB 57 NASNEPVE 64

RESULT 6
 Q7W5U6 PRELIMINARY; PRT; 114 AA.
 AC Q7W5U6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Putative exported protein.
 GN OrderedLocusNames=Bp3192;
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640447; CAE34086.1; -;
 DR InterPro; IPR008972; Cupredoxin.
 KW Complete proteome.
 SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;

Query Match 5.1%; Score 8; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 NASNEPVE 115
 |||||
 DB 57 NASNEPVE 64

RESULT 6
 Q7W5U6 PRELIMINARY; PRT; 114 AA.
 AC Q7W5U6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Putative exported protein.
 GN OrderedLocusNames=Bp3192;
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640447; CAE34086.1; -;
 DR InterPro; IPR008972; Cupredoxin.
 KW Complete proteome.
 SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;

ID Q7W5U6 PRELIMINARY; PRT; 114 AA.
 AC Q7W5U6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative exported protein.
 GN OrderedLocusNames=Bp3192;
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=12822 / ATCC BAA-587;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640432; CAE38477.1; -;
 DR InterPro; IPR008972; Cupredoxin.
 KW Complete proteome.
 SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;

Query Match 5.1%; Score 8; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 NASNEPVE 115
 |||||
 DB 57 NASNEPVE 64

RESULT 7
 Q7W5U7 PRELIMINARY; PRT; 114 AA.
 AC Q7W5U7;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative exported protein.
 GN OrderedLocusNames=BB3592;
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640447; CAE34086.1; -;
 DR InterPro; IPR008972; Cupredoxin.
 KW Complete proteome.
 SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;

Query Match 5.1%; Score 8; DB 2; Length 114;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 NASNEPVE 115
 |||||
 DB 57 NASNEPVE 64

RESULT 6
 Q7W5U6 PRELIMINARY; PRT; 114 AA.
 AC Q7W5U6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Putative exported protein.
 GN OrderedLocusNames=Bp3192;
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 ON NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feitwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640447; CAE34086.1; -;
 DR InterPro; IPR008972; Cupredoxin.
 KW Complete proteome.
 SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;

DR InterPro; IPR008972; Cupredoxin.
KW Complete proteome. 114 AA; 12335 MW; 731F7814735BA413 CRC64;
SQ SEQUENCE 114 AA; 12335 MW; 731F7814735BA413 CRC64;

Query Match 5.1%; Score 8; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 NASNEPVE 115
Db 57 NASNEPVE 64
|||||||

RESULT 8
ODO2_BUCAI STANDARD; PRT; 420 AA.
ID ODO2_BUCAI
AC P57389;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex (EC 2.3.1.61) (E2)
DE (Dihydrolipoamide succinyltransferase component of 2-oxoglutarate dehydrogenase complex).
DE Name=sucB; OrderedLocusName=BU303;
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077; DOI=10.1038/35024074;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: The 2-oxoglutarate dehydrogenase complex catalyzes the overall conversion of 2-oxoglutarate to succinyl-CoA and CO(2). It contains multiple copies of three enzymatic components: 2-oxoglutarate dehydrogenase (E1), dihydrolipoamide succinyltransferase (E2) and lipoamide dehydrogenase (E3) (By similarity).
CC -!- CATALYTIC ACTIVITY: Succinyl-CoA + enzyme N(6) - (dihydrolipoyl)lysine = CoA + enzyme N(6) - (S-succinyl)dihydrolipoyllysine.
CC -!- COFACTOR: Contains 1 covalently bound lipoyl cofactor (By similarity).
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SUBUNIT: Forms a 24-polypeptide structural core with octahedral symmetry (By similarity).
CC -!- SIMILARITY: Belongs to the 2-oxoacid dehydrogenase family.
CC -!- SIMILARITY: Contains 1 lipoyl-binding domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF001118; BAB13012.1; -.
CC HSP; P07016; 1C4T.
DR InterPro; IPR001078; 2Oxoacid dh.
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR011053; Hybrid_motif.
DR InterPro; IPR003016; Lipoyl_BS.
DR InterPro; IPR006255; SucB.
DR Pfam; PF00198; 2-oxoacid dh; 1.
DR Pfam; PF00364; Biotin lipoyl; 1.
DR ProDom; PD001115; 2Oxoacid_dh; 1.

TIGRFAMs; TIGR01347; sucB; 1.
DR PROSITE; PS00968; BIOTINYL_LIPOYL; 1.
DR PROSITE; PS00189; LIPOYL; FALSE_NEG.
KW Acyltransferase; Complete proteome; Lipoyl; Transferase;
KW Tricarboxylic acid cycle.
FT DOMAIN 1 77 Biotinyl/lipoyl.
FT BINDING 44 44 Lipoyl (potential).
FT ACT_SITE 391 391 Potential.
FT ACT_SITE 395 395 Potential.
SQ SEQUENCE 420 AA; 48091 MW; E0028D647A5CE34C CRC64;

Query Match 5.1%; Score 8; DB 1; Length 420;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 EKEGKVI 52
Db 62 EKEGKVI 69
|||||||

RESULT 9
Q26001 PRELIMINARY; PRT; 459 AA.
ID Q26001
AC Q26001;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE RabGDI protein.
DE Name=rabGDI;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=97001683; PubMed=8844675; DOI=10.1016/0166-6851(96)02606-0;
RA Attal G., Langsley G.;
RT "A Plasmodium falciparum homologue of a rab specific GDP dissociation inhibitor.";
RL Mol. Biochem. Parasitol. 79:91-95(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Langsley G. W.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X93166; CAA63653.1; -.
DR HSP; P21856; 1DST.
DR GO; GO:0005093; F:RAB GDP-dissociation inhibitor activity; IEA.
DR GO; GO:0015031; F:protein transport; IEA.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR000806; RabGDI.
DR InterPro; IPR002005; Rab_GDI_REP.
DR Pfam; PF00996; GDI; 1.
DR PRINTS; PR00892; RABGDI.
DR PRINTS; PR00891; RABGDIREP.
SQ SEQUENCE 459 AA; 52311 MW; DE1CCA47B124D8EA CRC64;

Query Match 5.1%; Score 8; DB 2; Length 459;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTLERIKL 68
Db 212 LTLERIKL 219
|||||||

RESULT 10
Q81501 PRELIMINARY; PRT; 459 AA.
ID Q81501
AC Q81501;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE RabGDI protein.
GN ORFNames=PF02060c;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum";
RL Nature 419:498-511(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Hyman R.W., Fung E., Conway A., Kurdi O., Mao J., Miranda M.,
RA Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS014851; AAN36496.1; -;
DR HSSP; P21856; 1D5T.
DR GO; GO:0005093; F:RAB GDP-dissociation inhibitor activity; IEA.
DR GO; GO:0015031; P:protein transport; IEA.
DR GO; GO:0041087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR008066; RabGDI.
DR InterPro; IPR020005; Rab_GDI_REP.
DR Pfam; PF00996; GDI; 1.
DR PRINTS; PR00892; RABGDI.
DR PRINTS; PR00891; RABGDIREP.
SQ SEQUENCE 459 AA; 52339 MW; 33AA113323679C2 CRC64;
Query Match 5.1%; Score 8; DB 2; Length 459;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 61 LTLEIRKL 68
DB 212 LTLEIRKL 219
|||||
RESULT 11
ID Q8UBB2 PRELIMINARY; PRT; 503 AA.
AC Q8UBB2; Q7CRL8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Alpha-L-arabinofuranosidase (AGR_L_3408p).
GN OrderedLocNames=AGR_L_3408, AtcU3104;
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Dupont;
RX MEDLINE=21608550; PubMed=11743193; DOI=10.1126/science.1066804;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;

RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58.";
RL Science 294:2317-2323 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Cereon;
RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Houmello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Qurollo K., Gordon J., Vaudin M., Tarchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Planagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58";
RL Science 294:2323-2328 (2001).
DR EMBL; AF009241; AAL43920.1; -;
DR EMBL; AF008372; AAK90279.1; -;
DR PIR; AB2938; AB2938.
DR PIR; E98344; E98344.
DR HSSP; Q9XBQ3; 1OW9.
DR InterPro; IPR010720; Alpha-L-AF_C.
DR Pfam; PF06964; Alpha-L-AF_C; 1.
KW Complete proteome.
SQ SEQUENCE 503 AA; 56495 MW; 3514F976D1854748 CRC64;
Query Match 5.1%; Score 8; DB 2; Length 503;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 25 AGYDLRVG 32
DB 455 AGYDLRVG 462
|||||
RESULT 12
ID Q9UQN9 PRELIMINARY; PRT; 34 AA.
AC Q9UQN9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE LSFR3A protein (Fragment).
GN Name=LSFR3A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99299247; PubMed=10369878; DOI=10.1093/hmg/8.7.1313;
RA Gilley J., Fried M.;
RT "Extensive gene order differences within regions of conserved synteny
between the Fugu and human genomes: implications for chromosomal
evolution and the cloning of disease genes";
RL Hum. Mol. Genet. 8:1313-1320 (1999).
DR EMBL; Y17458; CAB44351.1; -;
FT NON_TER 1 1
FT NON_TER 34 34
SQ SEQUENCE 34 AA; 4023 MW; 1F90377C28FA152D CRC64;
Query Match 4.5%; Score 7; DB 2; Length 34;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 EAFVKGK 40
DB 26 EAFVKGK 32
|||||
RESULT 13
ID Q9XQ05 PRELIMINARY; PRT; 37 AA.
QY 37 EAFVKGK 40
DB 26 EAFVKGK 32
|||||

AC Q9XQ05;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Ribosomal protein L36.
 OS Toxoplasma gondii.
 OG Chloroplast.
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;
 OC Toxoplasma.
 OX NCBI_TaxID=5811;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kissinger J.C., Donald R.G., Moulton A.L., Guttell R., Aiello D.P.,
 RA Lang-Unaach N., Roos D.S.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Belongs to the L36P family of ribosomal proteins.
 DR EMBL; U87145; AAD41142.1; -.
 DR GO; GO:0009507; C:chloroplast; IEA.
 DR GO; GO:0005840; C:ribosome; IEA.
 DR GO; GO:0003735; F:protein biosynthesis; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR000473; Ribosomal_L36.
 DR Pfam; PF00444; Ribosomal_L36; 1.
 DR TIGRFAMs; TIGR01022; rpmJ_bact; 1.
 KW Chloroplast; Ribonucleoprotein; Ribosomal protein.
 SQ SEQUENCE 37 AA; 4546 MW; EDD8C59BEEB5E6AD CRC64;
 Query Match 4.5%; Score 7; DB 2; Length 37;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 76 MKIRSSL 82
 DB 1 MKIRSSL 7

RESULT 14
 Q7P2B1 PRELIMINARY; PRT; 61 AA.
 AC Q7P2B1;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein.
 GN Name=FW0190;
 OS Fusobacterium nucleatum subsp. vincentii ATCC 49256.
 OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
 OC Fusobacterium.
 OX NCBI_TaxID=209882;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 49256;
 RA Karpatal V., Ivanova N., Anderson I., Reznik G., Bhattacharyya A.,
 RA Gardner W.L., Mikhailova N., Larsen N., D'Souza M., Walunas T.,
 RA Haselkorn R., Overbeek R., Kyrpides N.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 preliminary data.
 DR EMBL; AABF01000162; EAA23292.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 61 AA; 7335 MW; AF4936E8683E2348 CRC64;

Query Match 4.5%; Score 7; DB 2; Length 61;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 17 FSEESLQ 23
 DB 9 FSEESLQ 15

RESULT 15

SCX4 MESMA
 ID SCX4 MESMA STANDARD; PRT; 64 AA.
 AC P58328;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Alpha-like neurotoxin Bmk-IV (Bmk4) (Bmk M4) (BmkM4).
 OS Mesobuthus martensii (Machurian scorpion) (Buthus martensii).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
 OC Buthida; Buthoidea; Buthidae; Mesobuthus.
 OX NCBI_TaxID=34649;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Venom;
 RX MEDLINE=97346880; PubMed=9203297; DOI=10.1016/S0041-0101(96)00146-8;
 RA Luo M.-J., Xiong Y.-M., Wang M., Wang D.-C., Chi C.-W.;
 RL "Purification and sequence determination of a new neutral mammalian
 neurotoxin from the scorpion Buthus martensii Karsch.";
 RL Toxicon 35:723-731(1997).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.3 ANGSTROMS).
 RX MEDLINE=99425298; PubMed=10493862; DOI=10.1006/jmbi.1999.3036;
 RA He X.-L., Li H.-M., Zeng Z.-H., Liu X.-Q., Wang M., Wang D.-C.;
 RL "Crystal structures of two alpha-like scorpion toxins: non-proline cis
 peptide bonds and implications for new binding site selectivity on the
 sodium channel.";
 RL J. Mol. Biol. 292:125-135(1999).
 CC -!- FUNCTION: Binds to sodium channels and inhibits the inactivation
 of the activated channels, thereby blocking neuronal transmission.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.
 CC -!- TOXIC DOSE: LD(50) is 4 mg/kg by intravenous injection into mice.
 CC -!- SIMILARITY: Belongs to the alpha/beta-scorpion toxin family.
 CC Alpha-toxin subfamily. Alpha-like toxin subclase.
 PDB; 1SN4; X-ray; A-
 DR InterPro; IPR003614; Knot1.
 DR InterPro; IPR001219; Neurotoxin.
 DR InterPro; IPR002061; Scorpion_toxinL.
 DR Pfam; PF00537; Toxin_3; 1.
 DR PRINTS; PR00284; SCORPNTOXIN.
 DR PRINTS; PR00285; SCORPNTOXIN.
 DR PRODOM; PD000908; Scorpion_toxinL; 1.
 DR SMART; SM00505; Knot1; 1.
 KW 3D-structure; Direct protein sequencing; Ionic channel inhibitor;
 KW Neurotoxin; Sodium channel inhibitor; Toxin.
 FT DISULFID 12 63
 FT DISULFID 16 36
 FT DISULFID 22 46
 FT DISULFID 26 48
 FT STRAND 2 8
 FT TURN 9 11
 FT STRAND 12 13
 FT STRAND 15 15
 FT STRAND 19 28
 FT TURN 29 30
 FT STRAND 33 40
 FT TURN 41 42
 FT STRAND 43 51
 FT TURN 53 54
 FT STRAND 57 57
 SQ SEQUENCE 64 AA; 7021 MW; 20C7427C1C74DABB CRC64;

Query Match 4.5%; Score 7; DB 1; Length 64;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 66 IKLPDDV 72
 DB 49 IKLPDDV 55

Search completed: July 2, 2005, 12:38:13
 Job time : 32.9767 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:22:48 ; Search time 27.3953 Seconds
(without alignments)
1099.127 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 1998
Sequence: 1 MLHHVKLIYATSKRLVGKK.....KMKRELAERIWEIEKXLS 403

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1988	99.5	403	3	US-08-822-774-19
2	1988	99.5	403	3	US-09-632-711-19
3	1988	99.5	403	3	US-09-632-703B-19
4	1988	99.5	403	4	US-09-632-702-19
5	1988	99.5	403	4	US-09-399-003-19
6	635.5	31.8	437	3	US-08-822-774-51
7	635.5	31.8	437	3	US-09-632-711-51
8	635.5	31.8	437	3	US-09-632-703B-51
9	635.5	31.8	437	4	US-09-632-702-51
10	635.5	31.8	437	4	US-09-399-003-51
11	489	24.5	405	4	US-09-252-991A-24057
12	466	23.3	418	4	US-09-543-681A-7435
13	461.5	23.1	399	4	US-09-710-279-2576
14	461.5	23.1	400	3	US-09-134-001C-4785
15	458.5	22.9	431	4	US-09-540-236-3536
16	439	22.0	427	4	US-09-328-352-5205
17	438.5	21.9	444	3	US-08-822-774-52
18	438.5	21.9	444	3	US-09-632-711-52
19	438.5	21.9	444	3	US-09-632-703B-52
20	438.5	21.9	444	4	US-09-632-702-52
21	438.5	21.9	444	4	US-09-399-003-52
22	418.5	20.9	398	4	US-09-489-039A-10656
23	312.5	15.6	280	4	US-09-902-540-12020
24	262	13.1	183	4	US-09-583-110-4578
25	236	11.8	252	4	US-09-583-110-4579
26	235.5	11.8	245	4	US-09-107-532A-4448
27	235	11.8	188	4	US-09-107-532A-6818

28	233	11.7	262	4	US-09-134-000C-3905	Sequence 3905, Ap
29	219	11.0	154	4	US-09-107-433-4272	Sequence 4272, Ap
30	158	7.9	188	4	US-09-627-376-9	Sequence 9, Appl
31	158	7.9	188	4	US-10-047-676B-9	Sequence 9, Appl
32	151.5	7.6	181	2	US-08-645-193B-24	Sequence 24, Appl
33	149.5	7.5	181	2	US-08-392-625-22	Sequence 22, Appl
34	149.5	7.5	181	2	US-08-466-961A-22	Sequence 22, Appl
35	139	7.0	358	4	US-09-248-796A-15014	Sequence 15014, A
36	135	6.8	31	3	US-08-822-774-46	Sequence 46, Appl
37	135	6.8	31	3	US-09-632-711-46	Sequence 46, Appl
38	135	6.8	31	3	US-09-632-703B-46	Sequence 46, Appl
39	135	6.8	31	4	US-09-632-703-46	Sequence 46, Appl
40	135	6.8	31	4	US-09-399-003-46	Sequence 46, Appl
41	135	6.8	459	4	US-09-248-796A-19036	Sequence 19036, A
42	134	6.7	1500	3	US-09-323-472A-2	Sequence 2, Appl
43	134	6.7	1500	3	US-09-323-472A-4	Sequence 4, Appl
44	134	6.7	1500	3	US-09-323-472A-12	Sequence 12, Appl
45	134	6.7	1500	3	US-09-323-472A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-822-774-19
; Sequence 19, Application US/08922774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREPE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; NUMBER OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESS: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-822-774-19

Query Match 99.5%; Score 1988; DB 3; Length 403;
Best Local Similarity 100.0%; Pred. No. 9.4e-181;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLHHVKLIYATSKRLVGKKIVXXXPGSIAALDVKACEGLRHGAHVHWSAATKIIH 60
Db 1 MLHHVKLIYATSKRLVGKKIVXXXPGSIAALDVKACEGLRHGAHVHWSAATKIIH 60

QY 61 PYAWNLPCTGNPVITETGTFIEHVELAGHENKADLILVCPATANTISKIACGIDTTPVTT 120
DB 61 PYAWNLPCTGNPVITETGTFIEHVELAGHENKADLILVCPATANTISKIACGIDTTPVTT 120
QY 121 VTTTAPPHIPIMAPAMHETMYRHPIVRENIERLKKLGVEFTGPRIEGRKAVASIDEIV 180
DB 121 VTTTAPPHIPIMAPAMHETMYRHPIVRENIERLKKLGVEFTGPRIEGRKAVASIDEIV 180
QY 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFINASSGKMGVALAEADPRGAVTLIR 240
DB 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFINASSGKMGVALAEADPRGAVTLIR 240
QY 241 TKGSKVAFRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKASGKIKS 300
DB 241 TKGSKVAFRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKASGKIKS 300
QY 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKKLIBEGKQRIERAKADLVVNGT 360
DB 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKKLIBEGKQRIERAKADLVVNGT 360
QY 361 LEAFGSEENQVVLIGRDTFKLPKMKKRELAERIWDEIEKKLS 403
DB 361 LEAFGSEENQVVLIGRDTFKLPKMKKRELAERIWDEIEKKLS 403

RESULT 2

US-09-632-711-19
; Sequence 19, Application US/09632711
; Patent No. 633165
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,711
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/822,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-632-711-19

Query Match 99.5%; Score 1988; DB 3; Length 403;
Best Local Similarity 100.0%; Pred. No. 9.4e-181;

Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLHHVKLLIYATKSRKLVGKKIYVXXPGSIALDVKACGLRHGAHVHVMSEAAATKIHH 60
DB 1 MLHHVKLLIYATKSRKLVGKKIYVXXPGSIALDVKACGLRHGAHVHVMSEAAATKIHH 60
QY 61 PYAWNLPCTGNPVITETGTFIEHVELAGHENKADLILVCPATANTISKIACGIDTTPVTT 120
DB 61 PYAWNLPCTGNPVITETGTFIEHVELAGHENKADLILVCPATANTISKIACGIDTTPVTT 120
QY 121 VTTTAPPHIPIMAPAMHETMYRHPIVRENIERLKKLGVEFTGPRIEGRKAVASIDEIV 180
DB 121 VTTTAPPHIPIMAPAMHETMYRHPIVRENIERLKKLGVEFTGPRIEGRKAVASIDEIV 180
QY 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFINASSGKMGVALAEADPRGAVTLIR 240
DB 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRFINASSGKMGVALAEADPRGAVTLIR 240
QY 241 TKGSKVAFRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKASGKIKS 300
DB 241 TKGSKVAFRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKASGKIKS 300
QY 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKKLIBEGKQRIERAKADLVVNGT 360
DB 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKKLIBEGKQRIERAKADLVVNGT 360
QY 361 LEAFGSEENQVVLIGRDTFKLPKMKKRELAERIWDEIEKKLS 403
DB 361 LEAFGSEENQVVLIGRDTFKLPKMKKRELAERIWDEIEKKLS 403

RESULT 3

US-09-632-703B-19
; Sequence 19, Application US/09632703B
; Patent No. 6379553
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF
; Protein Complexes, Isolated PEF Proteins, and Methods for
; Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: <Unknown>
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,703B
; FILING DATE: 24-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BARKER, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 4121.0116-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 849-6613
; TELEFAX: (650) 849-6666
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO


```
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-632-703B-19

Query Match          99.5%; Score 1988; DB 3; Length 403;
Best Local Similarity 100.0%; Pred. No. 9.4e-181;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLHHVKLIYATSKRLVGKKIVXXXPGSTAALDVKACEGLIRHGAHVHVMSEAATKIIH 60
DB 1 MLHHVKLIYATSKRLVGKKIVXXXPGSTAALDVKACEGLIRHGAHVHVMSEAATKIIH 60

QY 61 PYAWNLPTGNPVITETITGFIEHVELAGEHENKADLLVCPATANTISKIACGIDDDTPVTT 120
DB 61 PYAWNLPTGNPVITETITGFIEHVELAGEHENKADLLVCPATANTISKIACGIDDDTPVTT 120

QY 121 VVTTAPPHIPIMIAPAMHETMYRHPIVRENIERLKKLGVEFGPIRIEGRKVASIDEIV 180
DB 121 VVTTAPPHIPIMIAPAMHETMYRHPIVRENIERLKKLGVEFGPIRIEGRKVASIDEIV 180

QY 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRITNASSGKMGVALAEADPRGAVTLIR 240
DB 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRITNASSGKMGVALAEADPRGAVTLIR 240

QY 241 TGSKVKAFRIRIKLVETVEEMLSAIENELSKKYDVVIMAAVSDFRPKIKAECKIKS 300
DB 241 TGSKVKAFRIRIKLVETVEEMLSAIENELSKKYDVVIMAAVSDFRPKIKAECKIKS 300

QY 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEYSKEKLIBEGKQIERAKADLVVGN 360
DB 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEYSKEKLIBEGKQIERAKADLVVGN 360

QY 361 LEAFGSEENQVVLIGRDFTKELPKMKKRELAERINWDEIEKXLS 403
DB 361 LEAFGSEENQVVLIGRDFTKELPKMKKRELAERINWDEIEKXLS 403

RESULT 4
US-09-632-702-19
; Sequence 19, Application US/09632702
; Patent No. 6444428
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; Lenahan, P.B.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,702
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 19:
```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-632-702-19

Query Match          99.5%; Score 1988; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 9.4e-181;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLHHVKLIYATSKRLVGKKIVXXXPGSTAALDVKACEGLIRHGAHVHVMSEAATKIIH 60
DB 1 MLHHVKLIYATSKRLVGKKIVXXXPGSTAALDVKACEGLIRHGAHVHVMSEAATKIIH 60

QY 61 PYAWNLPTGNPVITETITGFIEHVELAGEHENKADLLVCPATANTISKIACGIDDDTPVTT 120
DB 61 PYAWNLPTGNPVITETITGFIEHVELAGEHENKADLLVCPATANTISKIACGIDDDTPVTT 120

QY 121 VVTTAPPHIPIMIAPAMHETMYRHPIVRENIERLKKLGVEFGPIRIEGRKVASIDEIV 180
DB 121 VVTTAPPHIPIMIAPAMHETMYRHPIVRENIERLKKLGVEFGPIRIEGRKVASIDEIV 180

QY 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRITNASSGKMGVALAEADPRGAVTLIR 240
DB 181 YRVIKLHKKTLEGRKVLVTAGATREYIDPIRITNASSGKMGVALAEADPRGAVTLIR 240

QY 241 TGSKVKAFRIRIKLVETVEEMLSAIENELSKKYDVVIMAAVSDFRPKIKAECKIKS 300
DB 241 TGSKVKAFRIRIKLVETVEEMLSAIENELSKKYDVVIMAAVSDFRPKIKAECKIKS 300

QY 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEYSKEKLIBEGKQIERAKADLVVGN 360
DB 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEYSKEKLIBEGKQIERAKADLVVGN 360

QY 361 LEAFGSEENQVVLIGRDFTKELPKMKKRELAERINWDEIEKXLS 403
DB 361 LEAFGSEENQVVLIGRDFTKELPKMKKRELAERINWDEIEKXLS 403

RESULT 5
US-09-399-003-19
; Sequence 19, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Hogsafe, Holly
; APPLICANT: Hansen, Connie J
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
; TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying
; FILE REFERENCE: 4121.0116-02
; CURRENT APPLICATION NUMBER: US/09/399,003
; CURRENT FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/ US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(403)
; OTHER INFORMATION: "X" represents any amino acid
US-09-399-003-19

Query Match          99.5%; Score 1988; DB 4; Length 403;
```



```

; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Methanococcus Jannaschii
; SEQUENCE DESCRIPTION: SEQ ID NO: 51:
US-09-632-711-51

Query Match 31.8%; Score 635.5; DB 3; Length 437;
Best Local Similarity 37.9%; Pred. No. 6e-52;
Matches 162; Conservative 83; Mismatches 153; Indels 29; Gaps 12;

QY 1 MLHVKLIYATSKRLVGGKIVXXPGSIAALDV-KACEGLIRHGAHVHVMSEAATKII 59
DB 5 IMHPTKLLGKTSKLENKKILVAVTSSIAAETPKLMRELIRHGAHVYCIITEETKII 64

QY 60 HPYANMLPTGNPVITEITGF-----IEHVELAGEHENKADLILVCPATANTISKIACGID 114
DB 65 GKEALFSGCGNEVEYIEITGXXXXDIEHLLYXXXXNECDCLLIYPATANIISKINLGIA 124

QY 115 DPTVTVTTTAPPHIDIMTAPAMHETMYRHPVIRENIERLK-KLGVFEGIPREE----- 168
DB 125 DNVNTTALMFFGNKDFIFVPAMHENFNXAKIKRHIDKLEKDKIYIISPFEEXXXXX 184

QY 169 -GRAKVASIDEIVYRVIKKL-HKKTLEGRVLVTAGATREYIDPIRIFITNASSGKMGVAL 226
DB 185 XGKAKVANIEDVVKAVIEKIGNNLKKEGNRVLLNGTVEFDKVRVISNLSGKMGVAL 244

QY 227 ABEADPRG-AVTLIRTKGSKAFIRIKIKLVETVEEMLS-AIENELRKKYDVVIMAAA 284
DB 245 AEAFCKEGFYVEVITAMGLEPPYIK--NHKVLTAKEMLNKATEXXLXAKDFDIISAA 302

QY 285 VSDFRPKIAEGKIKS-----GRSITIELVPKPKIIDRIKEIQPNV-FLVGKAE--TSK 337
DB 303 ISDFTVESFEGKLSSEEXXXLILKRXNPKVLEELRRIYKDKXVIIGFKAENLDE 362

QY 338 EKLIIEGKQIERAKADLVVGNLTLEA---FGSEENQVVLIGRDFTKELPKMKKRELAER 393
DB 363 KELINRAKERLNKYNLNMIIANDLSKXHYFGDDYIEVYIITKYEVEKISGSKKXEISER 422

QY 394 IWDEIEK 400
DB 423 IVEKVKK 429

RESULT 8
US-09-632-703B-51
; Sequence 51, Application US/09632703B
; Patent No. 6379553
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: <Unknown>

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,703B
; FILING DATE: 24-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BARKER, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 4121.0116-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 849-6613
; TELEFAX: (650) 849-6666
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Methanococcus Jannaschii
; SEQUENCE DESCRIPTION: SEQ ID NO: 51:
US-09-632-703B-51

Query Match 31.8%; Score 635.5; DB 3; Length 437;
Best Local Similarity 37.9%; Pred. No. 6e-52;
Matches 162; Conservative 83; Mismatches 153; Indels 29; Gaps 12;

QY 1 MLHVKLIYATSKRLVGGKIVXXPGSIAALDV-KACEGLIRHGAHVHVMSEAATKII 59
DB 5 IMHPTKLLGKTSKLENKKILVAVTSSIAAETPKLMRELIRHGAHVYCIITEETKII 64

QY 60 HPYANMLPTGNPVITEITGF-----IEHVELAGEHENKADLILVCPATANTISKIACGID 114
DB 65 GKEALFSGCGNEVEYIEITGXXXXDIEHLLYXXXXNECDCLLIYPATANIISKINLGIA 124

QY 115 DPTVTVTTTAPPHIDIMTAPAMHETMYRHPVIRENIERLK-KLGVFEGIPREE----- 168
DB 125 DNVNTTALMFFGNKDFIFVPAMHENFNXAKIKRHIDKLEKDKIYIISPFEEXXXXX 184

QY 169 -GRAKVASIDEIVYRVIKKL-HKKTLEGRVLVTAGATREYIDPIRIFITNASSGKMGVAL 226
DB 185 XGKAKVANIEDVVKAVIEKIGNNLKKEGNRVLLNGTVEFDKVRVISNLSGKMGVAL 244

QY 227 ABEADPRG-AVTLIRTKGSKAFIRIKIKLVETVEEMLS-AIENELRKKYDVVIMAAA 284
DB 245 AEAFCKEGFYVEVITAMGLEPPYIK--NHKVLTAKEMLNKATEXXLXAKDFDIISAA 302

QY 285 VSDFRPKIAEGKIKS-----GRSITIELVPKPKIIDRIKEIQPNV-FLVGKAE--TSK 337
DB 303 ISDFTVESFEGKLSSEEXXXLILKRXNPKVLEELRRIYKDKXVIIGFKAENLDE 362

QY 338 EKLIIEGKQIERAKADLVVGNLTLEA---FGSEENQVVLIGRDFTKELPKMKKRELAER 393
DB 363 KELINRAKERLNKYNLNMIIANDLSKXHYFGDDYIEVYIITKYEVEKISGSKKXEISER 422

QY 394 IWDEIEK 400
DB 423 IVEKVKK 429

RESULT 9
US-09-632-702-51
; Sequence 51, Application US/09632702
; Patent No. 644428
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)

```


RESULT 11

US-09-252-991A-24057
; Sequence 24057, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24057
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24057

Query Match 24.5%; Score 489; DB 4; Length 405;
Best Local Similarity 34.8%; Pred. No. 4.7e-38;
Matches 136; Conservative 77; Mismatches 132; Indels 46; Gaps 16;

QY 14 RKLVGKIVXXPGSIAALDVKACEGLIR----HGAEVHVMSEAATKIHPYAMNPTG 69
DB 5 QRLYRKRIVLGVGGGIAA--YKSAE-LVRLRDQGAEEVRVMTQGGREFITPLTLQALSG 61
QY 70 NPVITEI-----TGFIEHVELAGEHENKADLLVCPATANTISKIACGIDDTPTVTVT 124
DB 62 HPVHTDLLDPAAEAAMGHIELA----RWADLVLIAPATADLMARLVQGVANDLLTTLVLA 117
QY 125 AFPHIPIMAPAMHETMYRHPVIRENIERKLKGVFEIGPRI-----BEGRAKVASIDE 178
DB 118 T--DAQIALAPAMQAMWRDTATQANELLRQGFHLFGPAAGSQAGDVGGLRMLEAE 175
QY 179 IVRVIKLHKHTLEGKRVLTAGATREYIDIRFITNASSGKMGVALAEADFRGA-VT 237
DB 176 LAQRAADCFORQALTGHVLTITAGTQENIDPVRYITNHSKGWGFALAEAAVEAGARVT 235
QY 238 LIRTKGSV---KAFRTRKIKLVETVEEMLSAIENELRSKYYDVVIMAAAVSDFRPKIKA 294
DB 236 LV--TGPVHLPTDRQVRV--DVVSARDMLAACEAEM---PCDLLIASAAVADYRPEVVA 288
QY 295 EGKIK---SGRSITTELVPXNPKIIDRIKEIQPNVFLVGFKAETSKELIESGKQIER 350
DB 289 AHLKKDPTSGEGLLIQLV-RNPDILATLAQRDRPFSVGFAAET--ENLLDYAARKIKD 345
QY 351 AKADLVVGNLTLE---AFGSEENQVVLIGRD 377
DB 346 KNLDLIVANDVANPSIGFNSDENAITVIDRD 376

RESULT 12

US-09-543-681A-7435
; Sequence 7435, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7435
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Proteus mirabilis

US-09-543-681A-7435

Query Match 23.3%; Score 466; DB 4; Length 418;
Best Local Similarity 34.2%; Pred. No. 7.7e-36;
Matches 149; Conservative 76; Mismatches 145; Indels 66; Gaps 20;
QY 5 VKL---IYATK---SRKLVGKIVXXPGSIAALDVKACEGLIRH-----CAEVHVMSEA 54
DB 3 VKLWTFIYAIRHFIMTTLHDKNIILGISGGIAA--YKRAPE-LVRLRDKGAIVRVVMTPA 59
QY 55 ATKIIHPYAWNLPNTGNPVITEI-----TGFIEHVELAGEHENKADLLVCPATANTISKI 109
DB 60 AHAFVTPLSIQAVSGFPVADLLDPAAEAAMGHIELG----KWADLILLAPATADLIARL 115
QY 110 ACIDIDTPVTVTVTAFPHIPIMAPAMHETMYRHPVIRENIERKLKGVFEITGP----- 164
DB 116 RMGMANDLLTTLCLAS--AAPAIAPAMNQMYRATITQENLTALAQRCGLIWGPDGSGQ 173
QY 165 ---RIEGR---AKVASIDEIVRVVVKLHKHTLEGKRVLTAGATREYIDIRFITNA 217
DB 174 ACDVGPGRMULDPLELVALAEQOFAI-----QHDFGKKITITAGPTREALDPVRFISNH 228
QY 218 SSGKMGVALAEADFRGA-VTLIR-----TKGSVKAFRIKIKLVETVEEMLSAIENE 270
DB 229 SSGKMGFATAQAQAQGAEVTLIAGPVTLPACVK-----RIDVESAQEMYQWMDI 281
QY 271 LRSKYYDVVIMAAVSDFRPKIKAEQKI-KSGRSITTELVPXNPKIIDRI-KBIQPNVFL 328
DB 282 AQSQ--DIFIGCAAVADYRAKQVAPEKIKKQGDVITIMI-KNPDIVASVGKMHVHRPFV 338
QY 329 VGFKAETSEKLEIEGKQIERAKADLVVGNLTLE---AFGSEENQVVLIGRDTKELPK 384
DB 339 VGFAAET--QNVVEYARRREKQKQLDICANDVSLKDGAFNSDNALHLIWAANGETRLPH 396
QY 385 MKKRELAERIWDIEIK 400
DB 397 SKTQLSHRLLDEIAK 412

RESULT 13

US-09-710-279-2576
; Sequence 2576, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; PRIOR FILING DATE: 2000-11-09
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2576
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-2576

Query Match 23.1%; Score 461.5; DB 4; Length 399;
Best Local Similarity 34.3%; Pred. No. 1.9e-35;
Matches 141; Conservative 75; Mismatches 148; Indels 47; Gaps 16;
QY 19 KKIIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKIHPYAWNLPNGPVT--- 74
DB 2 KHLIAVTTGGIAAYKAIDLTSLKIQSGYDVRVWLSDBAQEFVFTPLAFQAISRNPVYNTNF 61
QY 75 --EITGFIEHVELAGEHENKADLLVCPATANTISKIACGI-DDTPTVTVTVTAFPHIPI 131
DB 62 KEENPEEIQHVSLG-----DWADAIIVAPATANTIAKLSVGIADDLITLTLLATTPK--- 114

QY 132 MIAPAMHETMYRHPVIRENIEIRKLVGEFICP-----RIEGBRAKVASIDE 178
Db 115 FVAPAMNNVNNYNNPRTKNNKVLSDQGYFFIEPGSGYLACGVAKGRMEE-PMQLLSVIN 173
QY 179 IYRVIKKLLHKTLEGRVLTAGATREVIDIRFITNASSGKMGVALAEADFRGA-VT 237
Db 174 KEFTQKNVVKSSFSGKRALVTAGTVEIDPVRYVSNRSSGKMGYATAEARLDRGAIVT 233
QY 238 LIRTKGSVKAFRIRKIK-LKVTVEEMLSAINEIRSKKYDVVIMAAVSDRPRPKIKAE 296
Db 234 LI--SGPHTLSPEGINNVKVESADDMFOAVTE--RFAKQDIVIKAAVSDVTPMDILEH 289
QY 297 KIK---SGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKELIEBGRQIERAKA 353
Db 290 KLIKQEGGLSVQFK---RTKDILKYLGENKTHOYLVGFAAET--QNIQYALDKLKRKNA 344
QY 354 DLV----VGNLTLEAFGSEENQVVLIGRDTK-ELPKMKKRELAERIWDEIE 399
Db 345 DVIISNNVGDTSIGFSSDDNLTMHFKNNEKVNKKKKSALAHQIIEILE 395

RESULT 14
US-09-134-001C-4785 Query Match 23.1%; Score 461.5; DB 3; Length 400;
; Sequence 4785, Application US/09134001C Best Local Similarity 34.3%; Pred. No. 1.9e-35;
; Patent No. 6380370 Matches 141; Conservative 75; Mismatches 148; Indels 47; Gaps 16;
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007 EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4785
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4785

QY 19 KIVVXXPGSIALD-VKACEGLIRHGAHVAMSEAAATKIHPYAWNLPNGPVIT--- 74
Db 3 KHLIAVTGGIAAYKAIDUTSKLIQSGYDVRVWLSHQAQEFVTPLAFOAISRNPTVTNTF 62
QY 75 --EITGFIHVELAGEHENKADILVCPATANTISKIACGI-DDTPVTVTVTATPPHIPI 131
Db 63 KEENPEIQHVSIG---DWADAIIVAPATANTIAKLSVGIADLLITSLTATTPK--- 115
QY 132 MIAPAMHETMYRHPVIRENIEIRKLVGEFICP-----RIEGBRAKVASIDE 178
Db 116 FVAPAMNNVNNYNNPRTKNNKVLSDQGYFFIEPGSGYLACGVAKGRMEE-PMQLLSVIN 174
QY 179 IYRVIKKLLHKTLEGRVLTAGATREVIDIRFITNASSGKMGVALAEADFRGA-VT 237
Db 175 KEFTQKNVVKSSFSGKRALVTAGTVEIDPVRYVSNRSSGKMGYATAEARLDRGAIVT 234
QY 238 LIRTKGSVKAFRIRKIK-LKVTVEEMLSAINEIRSKKYDVVIMAAVSDRPRPKIKAE 296
Db 235 LI--SGPHTLSPEGINNVKVESADDMFOAVTE--RFAKQDIVIKAAVSDVTPMDILEH 290
QY 297 KIK---SGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKELIEBGRQIERAKA 353
Db 291 KLIKQEGGLSVQFK---RTKDILKYLGENKTHOYLVGFAAET--QNIQYALDKLKRKNA 345
QY 354 DLV----VGNLTLEAFGSEENQVVLIGRDTK-ELPKMKKRELAERIWDEIE 399

Db 346 DVIISNNVGDTSIGFSSDDNLTMHFKNNEKVNKKKKSALAHQIIEILE 396
RESULT 15
US-09-540-236-3536 Query Match 22.9%; Score 458.5; DB 4; Length 431;
; Sequence 3536, Application US/09540236 Best Local Similarity 33.0%; Pred. No. 4.2e-35;
; Patent No. 6673910 Matches 140; Conservative 88; Mismatches 147; Indels 49; Gaps 19;
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001 EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3536
; LENGTH: 431
; TYPE: PRT
; ORGANISM: M. catarrhalis
US-09-540-236-3536

QY 5 VKLIYATKSKLVGK--KIVVXXPGSIALD-DVKACEGLIRHGAHVAMSEAAATKIHP 61
Db 11 IMMFAKCOMKVMTOQLPKILLAVTGGIAAYKSAMLARLIKSGCQVRVMTAGACEFITP 70
QY 62 YAWNLPNTGNPVITEITG-----FIEHVELAGEHENKADILVCPATANTISKIACGIDT 116
Db 71 LTFQALTQOAVHTKLLDDEAERGMDHITLA---KWADLVVVAASANTIAKLAGGFADN 126
QY 117 PVTVTVTAFPHIPIMIAPIAMHETMYRHPVIRENIEIRKLVGEFICPRIEE-----GR 170
Db 127 LVTTVCLAT--DAPILIAPANMQMNAIVQDNLKRLQRFYGHIMTPDSGQACGDVGL 184
QY 171 AKVASIDEIVYRVI---KKLH-KKTLGKRVLTAGATREYIDPIRFINASSGKMGVAL 226
Db 185 GRLPEPEDLCOQILAFDKCRHIQPSLLGKTVVITAGATVEPIDPVRFLSNHSTGKMGYAL 244
QY 227 AEEADFRGA-VTLIRTKG-SVKA-FRIRKIKLVETVEEMLSAINEIRSKKYDVVIMAA 283
Db 245 ANACYHAGADVLVSGKHVSLKTPFGVR--RLDVGTAEQMLKVCLSV--CQKADIFIAAA 300
QY 284 AVSDRPRPKIKAEKGI---KSGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAET----- 335
Db 301 AVADFKAAASVADHKIKTKTKNQOTMTLQLI-KNPDVLATICTHTYPPDLICVGFAAETODANN 359
QY 336 -SKEKLIIEGKQIERAKADLVVGNLTLEAFGSEENQVVLIGRD---FTKELPKMKKREL 390
Db 360 CAKSKL---AAKQLDMIADVNSDKTI-GFGSENNAMTVFFAEQYDQMPQNLPKAHKSKI 415
QY 391 AERI 394
Db 416 AEQL 419

Search completed: July 2, 2005, 12:25:59
Job time : 29.3953 secs

GenCore version 5.1.1.6
Copyright (C) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 11:27:13 ; Search time 31.2558 Seconds
(without alignments)
2555.824 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPDWKIRKELIEPFSEE.....PYRGNVQSTRLAFSKRRKL 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	ID	Description
1	806	100.0	156	1	DCD_PYRFU	Q8x251 pyrococcus
2	806	100.0	156	2	Q6T180	Q6t180 pyrococcus
3	768	95.3	156	1	DCD_PYRAB	O57706 pyrococcus
4	703	87.2	154	1	DCD_PYRAB	Q9uxs8 pyrococcus
5	257.5	31.9	166	1	DCD_TTEAC	Q9hkk0 thermoplasma
6	245	30.4	172	2	Q684E1	Q684e1 sulfolobus
7	244	30.3	158	2	Q8V9P5	Q8v9p5 sulfolobus
8	243	30.1	180	1	DCD_AQUAE	O67539 aquifex ae
9	242.5	30.1	177	1	DCD_BACHD	Q9kfv3 bacillus ha
10	242	30.0	173	1	DCD_ACTAM	Q02103 acidianus a
11	237	29.4	158	2	O71028	O71028 archaeal vi
12	237	29.4	158	2	O777M3	O777m3 sulfolobus
13	236	29.3	176	1	DCD_PYRAE	Q8zw23 pyrobaculum
14	233	28.9	181	2	Q72G14	Q72g14 thermus the
15	233	28.9	183	1	DCD_SULTO	Q976g3 sulfolobus
16	232.5	28.8	181	1	DCD_ABRPE	Q9vfa8 aeropyrum p
17	232	28.8	143	2	Q612T7	Q612t7 picophilus
18	227.5	28.2	198	2	Q74MA7	Q74ma7 nanoarchaeu
19	223.5	27.7	157	2	Q97B24	Q97b24 thermoplas
20	222.5	27.6	190	2	Q73T98	Q73t98 methanopyru
21	220	27.3	193	1	DCD_METKA	Q8tyk5 methanopyru
22	218	27.0	189	1	DCD_CORGL	Q8nlt9 corynebacte
23	216	26.8	191	1	DCD_STRCO	Q9x8w0 streptomyce
24	214	26.6	191	1	DCD_STRAW	Q8acz9 streptomyce
25	208	25.8	187	2	Q6NEW7	Q6new7 corynebacte
26	207.5	25.7	184	1	DCD_SULSO	Q980t7 sulfolobus
27	206	25.6	193	1	DCD_COREF	Q8fm44 corynebacte
28	205.5	25.5	196	1	DCD_TROW8	Q83h71 tropheryma
29	205.5	25.5	196	1	DCD_TROWT	Q820x8 tropheryma
30	203.5	25.2	201	2	Q6AC71	Q6ac71 leifsonia x
31	199	24.7	197	1	DCD_METTH	O27875 methanobact

Query Match 100.0%; Score 806; DB 1; Length 156;

Best Local Similarity 100.0%; Pred. No. 7.6e-67;

Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

32	198	24.6	190	1	DCD_MYCBO	Q7u297 mycobacteri
33	198	24.6	190	1	DCD_MYCTU	Q07247 mycobacteri
34	197.5	24.5	190	1	DCD_MYCLE	Q9cb17 mycobacteri
35	191.5	23.8	193	1	DCD_BIFLO	Q89478 bifidobacte
36	189.5	23.5	186	1	DCD_CAMJE	Q9pn07 campylobact
37	185	23.0	193	1	DCD_BUCBP	P59464 buchnera ap
38	184.5	22.9	195	1	DCD_HALN1	Q9hsq3 halobacteri
39	181	22.5	188	1	DCD_HELPY	O25136 helicobacte
40	180	22.3	188	1	DCD_HELPU	Q9zkd0 helicobacte
41	179.5	22.3	188	1	DCD_HELHP	Q7vj1q helicobacte
42	177	22.0	193	1	DCD_YERPE	Q8zfk8 versinia pe
43	177	22.0	193	2	Q66C69	Q66c69 versinia ps
44	174	21.6	193	1	DCD_BUCAI	P57209 buchnera ap
45	172.5	21.4	189	2	Q6FE29	Q6fe29 acinetobact

ALIGNMENTS

RESULT 1

ID DCD_PYRFU STANDARD; PRT; 156 AA.
AC Q8X251;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP deaminase).
DE deaminase).
GN Name=dcid; OrderedLocusNames=PF1996;
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21664347; PubMed=11782527; DOI=10.1073/pnas.012372799;
RA Hogrefe H.H., Hansen C.J., Scott B.R., Nielson K.B.;
RT "Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation."
RL Proc. Natl. Acad. Sci. U.S.A. 99:596-601(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY056005; AAL47572.1; -;
DR EMBL; AE010292; AAL02120.1; -;
DR HSSP; Q57872; IOGH.
DR HAMAP; MF_00146; -; 1.
DR InterPro; IPR003232; dCTP deaminase.
DR InterPro; IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
DR Complete proteome; Hydrolyase; Nucleotide metabolism.
SQ SEQUENCE 156 AA; 17869 MW; 00D57A5A0207B451 CRC64;

```

QY 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDIVKEGKVIPPPEYALI 60
Db 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDIVKEGKVIPPPEYALI 60

QY 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSFAPWDPGWDGNTLTLMLYNASNEPVELRYGE 120
Db 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSFAPWDPGWDGNTLTLMLYNASNEPVELRYGE 120

QY 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156
Db 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156

RESULT 2
ID Q6T180 PRELIMINARY; PRT; 156 AA.
AC Q6T180;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE DUTPase.
OS Pyrococcus woesei.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX Pyrococcus.
OC NCBI_TaxID=2262;
RN [1]

SEQUENCE FROM N.A.
RX MEDLINE=22844507; PubMed=12963343; DOI=10.1016/S1046-5928(03)00108-6;
RA Dabrowski S., Klier Ahring B.;
RT "Cloning, expression, and purification of the His6-tagged hyper-
RT thermostable dUTPase from Pyrococcus woesei in Escherichia coli:
RT application in PCR.";
RL Protein Expr. Purif. 31:72-78(2003).
DR EMBL; AY443493; AARI5897.1; -.
DR GO; GO:0046080; P:dutp metabolism; IEA.
DR InterPro; IPR003232; dCTP deaminase.
DR InterPro; IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
SQ SEQUENCE 156 AA; 17869 MW; 00D57A5A0207B451 CRC64;

Query Match 100.0%; Score 806; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 7.6e-67;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDIVKEGKVIPPPEYALI 60
Db 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDIVKEGKVIPPPEYALI 60

QY 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSFAPWDPGWDGNTLTLMLYNASNEPVELRYGE 120
Db 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSFAPWDPGWDGNTLTLMLYNASNEPVELRYGE 120

QY 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156
Db 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156

RESULT 3
ID DCD_PYRHO STANDARD; PRT; 156 AA.
AC O57706;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP
DE deaminase).
GN Name=dcdd; OrderedLocusNames=PH1997;
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX Pyrococcus.
OC NCBI_TaxID=53953;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Koeugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AP000007; BAA31124.1; -.
DR PIR; E71216; E71216.
DR HSSP; Q57872; IOGH.
DR HAMAP; MF_00146; -.
DR InterPro; IPR003232; dCTP deaminase.
DR InterPro; IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
KW Complete proteome; Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 156 AA; 17871 MW; 96B2C2C5039A985 CRC64;

Query Match 95.3%; Score 768; DB 1; Length 156;
Best Local Similarity 93.6%; Pred. No. 2.6e-63;
Matches 146; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDIVKEGKVIPPPEYALI 60
Db 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAFVKGKLDIVKEGKVIPPPEYALI 60

QY 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSFAPWDPGWDGNTLTLMLYNASNEPVELRYGE 120
Db 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSFAPWDPGWDGNTLTLMLYNASNEPVELRYGE 120

QY 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156
Db 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKKL 156

RESULT 4
ID DCD_PYRAB STANDARD; PRT; 154 AA.
AC Q9UKS8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP
DE deaminase).
GN Name=dcdd; OrderedLocusNames=PYRAB17800; ORFNames=PAB1164;
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RX MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;

```


RT "An integrated analysis of the genome of the hyperthermophilic
archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entireties requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ248288; CAB50685.1; -;
DR PIR; G75030; G75030.
DR HSP; Q57872; LOGH.
DR HAMAP; MF_00146; -; 1.
DR InterPro; IPR003232; dCTP deaminase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
KW Complete proteome; Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 154 AA; 17758 MW; C031BEE415094DDB CRC64;

Query Match 87.2%; Score 703; DB 1; Length 154;
Best Local Similarity 85.6%; Pred. No. 2.7e-57;
Matches 131; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

QY 1 MLLPDWKIRKEILIEPFSESLQAGYDLRVGREAFAVKGKLDIVEGKVVPPREYALI 60
DB 1 MLLPDWKIRKEILIEPFSESLQAGYDLRVGREAFAVKGKLDIVEGKVVPPREYALI 60

QY 61 LTLERIKLPDDVGMGDMKIRSLAREGVIGSFPAWDPGWDGNLTMLYNASNEPVELRYGE 120
DB 61 LTLERIKLPDDVGMGDMKIRSLAREGVIGSFPAWDPGWDGNLTMLYNASNEPVELRYGE 120

QY 121 RFVQIAFLIEGPARNPYRGNGQSTRFLAFSKR 153
DB 121 RFVQIAFLIEGPARNPYRGNGQSTRFLAFSKR 153

RESULT 5
DCD THEAC STANDARD; PRT; 166 AA.
AC Q9HKK0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP
deaminase).
GN Name=cdcd; orderedLocusNames=Ta0598;
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]_TaxID=2303;
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001; DOI=10.1038/35035069;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
acidophilum.";
RL Nature 407:508-513(2000).
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entireties requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL445064; CAC11737.1; ALT_INIT.
DR HSP; Q07199; 1MQ7.
DR HAMAP; MF_00146; -; 1.
DR InterPro; IPR003232; dCTP deaminase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
KW Complete proteome; Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 166 AA; 18293 MW; 1F4C9467CA3BFBD8 CRC64;

Query Match 31.9%; Score 257.5; DB 1; Length 166;
Best Local Similarity 41.3%; Pred. No. 6.5e-16;
Matches 64; Conservative 22; Mismatches 50; Indels 19; Gaps 5;

QY 1 MLLPDWKIRKEILIEPFSESLQAGYDLRVGREAFAVKGKLDIVEGKVVIP 53
DB 1 MLLPDWKIRKEILIEPFSESLQAGYDLRVGREAFAVKGKLDIVEGKVVIP 53

QY 54 PREYALITLERIKLPDDVGMGDMKIRSLAREGVIGSFPAWDPGWDGNLTMLYNASNEP 113
DB 54 PREYALITLERIKLPDDVGMGDMKIRSLAREGVIGSFPAWDPGWDGNLTMLYNASNEP 113

QY 114 VELRYGERFVQIAFLIEGPARNPYR---GNYQGS 145
DB 114 VELRYGERFVQIAFLIEGPARNPYR---GNYQGS 145

QY 112 VNLRRGERIAQIVFVKMIGSAEKPYHRSNGVQNS 146
DB 112 VNLRRGERIAQIVFVKMIGSAEKPYHRSNGVQNS 146

RESULT 6
Q684E1 PRELIMINARY; PRT; 172 AA.
ID Q684E1;
AC Q684E1;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE dUTPase.
GN Name=dut;
OS Sulfolobus tengchongensis spindle-shaped virus STSV1.
OC Viruses; dsDNA viruses, no RNA stage; Fuselloviridae; Fusellovirus.
OX NCBI_TaxID=285013;
RN [1]
RP SEQUENCE FROM N.A.
RA Xiang X., Chen L., Huang X., She Q., Huang L.;
RT "The Sulfolobus tengchongensis spindle-shaped virus STSV1: virus-host
interactions and genomic features.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ783769; CAH04216.1; -;
DR InterPro; IPR003232; dCTP deaminase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP deaminase; 1.
SQ SEQUENCE 172 AA; 19974 MW; 0999F5F1488F095 CRC64;

Query Match 30.4%; Score 245; DB 2; Length 172;
Best Local Similarity 37.9%; Pred. No. 9.8e-15;
Matches 55; Conservative 28; Mismatches 48; Indels 14; Gaps 4;

QY 12 ILIEPFSESLQAGYDLRVGRE--AFVKGKLD-----VEKE--GKVVIPPREYA 58
DB 17 IKIEPLREDTIRENGVGLRIGNEIVRFKQKRFDPDKSDIDDFIEKEGVNEFIINPHEHV 76

QY 59 LTLERIKLPDDVGMGDMKIRSLAREGVIGSFPAWDPGWDGNLTMLYNASNEPVELRY 118
DB 77 LLVTEYVKLPNDVMAFVNLRSFTRFARLGFIPPTIVDAGFEGQLTIELV-GSEFPKLY 135

QY 119 GERFVQIAFLIEGPARNPYRGNYQ 143
DB 136 GMRFIHLIAFKLTLPVERPYNGYQ 160

DR HAMAP: MF 00146; -, 1.
DR InterPro: IPR003232; dCTP deaminase.
DR InterPro: IPR008180; DeoxyUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP deaminase; 1.
DR Complete proteome; Hydrolyase; Nucleotide metabolism.
KW SEQUENCE 177 AA; 19897 MW; D15AE7538784782B CRC64;

Query Match 30.1%; Score 242.5; DB 1; Length 177;
Best Local Similarity 37.2%; Pred. No. 1.7e-14;
Matches 58; Conservative 30; Mismatches 47; Indels 21; Gaps 6;

QY 10 KEILIEPFSESLQAGYDLRVGREAFV-----KGKLIDVEK-----EGKVIVP 54
DB 15 KEILIEPFSESLQAGYDLRVGREAFV-----KGKLIDVEK-----EGKVIVP 73

QY 55 REYALITLERIKLPDDVGMGDKIRSLAREGV-IGSFADVDPGWDGNLTMLYNASNEP 113
DB 74 HTPFLATTMETVKLPNHLATFVGRSSVGLGLFIQAGNAGVDPGFNGQITLFLFNARLP 133

QY 114 VELRYGERFVQIAFIRLEGPARNPYRGNY---QGST 146
DB 134 IELPIGRICQLVFAEVTGEVA-PYQGYLFQKGAT 168

RESULT 10
DCD ACIAM STANDARD; PRT; 173 AA.
AC Q02103;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP deaminase).
GN Name=dcd;
OS Acidianus ambivalens (Desulfohalobium ambivalens).
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Acidianus.
OX NCBI_TaxID=2283;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lei 10 / DSM 3772;
RX MEDLINE=93065206; PubMed=14375556;
RA Kletzin A.;
RT "Molecular characterisation of a DNA ligase gene of the extremely thermophilic archaeon Desulfohalobium ambivalens shows close phylogenetic relationship to eukaryotic ligases.";
RL Nucleic Acids Res. 20:5389-5396(1992).
RN [2]
RP SIMILARITY.
RX MEDLINE=95206934; PubMed=7899076;
RA Ouzounis C., Kyriades N., Sander C.;
RT "Novel protein families in archaean genomes.";
RL Nucleic Acids Res. 23:565-570(1995).
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X63438; CAA45033.1; -.
DR PIR; S26382; S26382.
DR HSSP; Q57872; IOGH.
DR HAMAP; MF 00146; -, 1.
DR InterPro: IPR003232; dCTP deaminase.
DR InterPro: IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.

DR ProDom: PD004900; dCTP deaminase; 1.
KW SEQUENCE 173 AA; 19858 MW; B4D922503CD4B25A CRC64;

Query Match 30.0%; Score 242; DB 1; Length 173;
Best Local Similarity 35.9%; Pred. No. 1.9e-14;
Matches 56; Conservative 30; Mismatches 56; Indels 14; Gaps 3;

QY 12 ILIEPFSESLQAGYDLRVG-----RFAFVKGK-----LIDVEKGVVPPREYA 58
DB 17 IVISPLTQDTIRENGVDLRVGGGEIARFKTKTDEIYEDGKDRSPFYETIEKGEDEFIYPNEHV 76

QY 59 LILTLERIKLPDDVGMGDKIRSLAREGVIGSFADVDPGWDGNLTMLYNASNEPVELRY 118
DB 77 LIVTEYVKLPDDVMAFVNLRSFARLGLFVPTIVDAGFEGQLTIEVL-GSAPPVKIKR 135

QY 119 GERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRK 154
DB 136 GTRFHLHIFARTLTPVENPYHGKYQGQGVTLPPKPK 171

RESULT 11
O71028 PRELIMINARY; PRT; 158 AA.
AC O71028;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE dUTPase (EC 3.6.1.23).
OS Archaeal virus SIRV.
OC Viruses; unclassified viruses.
OX NCBI_TaxID=66287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98165772; PubMed=9497317; DOI=10.1074/jbc.273.11.6024;
RA Prangishvili D., Klenk H.P., Jakobs G., Schmiedchen A., Hanselmann C., Holz I., Zillig W.;
RT "Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV.";
RL J. Biol. Chem. 273:6024-6029(1998).
DR EMBL; AF022221; AAC15873.1; -.
DR HSSP; Q57872; 1PKK.
DR GO; GO:0004170; F:dUTP diphosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0046080; P:dUTP metabolism; IEA.
DR InterPro: IPR003232; dCTP deaminase.
DR InterPro: IPR008180; DeoxyUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP deaminase; 1.
KW Hydrolyase.
SQ SEQUENCE 158 AA; 17906 MW; 5BA47765E0889190 CRC64;

Query Match 29.4%; Score 237; DB 2; Length 158;
Best Local Similarity 34.4%; Pred. No. 4.9e-14;
Matches 54; Conservative 36; Mismatches 61; Indels 6; Gaps 2;

QY 1 MLLPDWKIR-----KEILIEPFSESLQAGYDLRVGREAFVKGKLIDVEKGVVPPR 55
DB 1 MILSDRDIKTYNSKKLVINPLSEDTIRENGVDLKGNEIVRIKENMEKEVGDEFIYPN 60

QY 56 EVALITLERIKLPDDVGMGDKIRSLAREGVIGSFADVDPGWDGNLTMLYNASNEPVE 115
DB 61 EHVLTITKEYIKLSNDIAFACNLRLSTFARKGLLIPTIVDAGFEGQLTIEVL-GSSIIPVK 119

QY 116 LAYGRFVQIAFIRLEGPARNPYRGNYQGSTRLAFSK 152
DB 120 LKSGERFHLHIFARTLTPVEKPYNGKYQKQKGVTLAK 156

RESULT 12
Q777W3 PRELIMINARY; PRT; 158 AA.
ID Q777W3
AC Q777W3;

Db 61 KNFDPKSFVEEGEEVWIPPNFALTRSLRYIRMPDNVIAIALGKSTYARCGIUVNVTPL 120
QY 95 DFGWDGNTLMLYNASNEPVELRYGERFVQIAFIRLEGPARN-PYR--GNYQG 144
Db 121 EPGWEGHTLEISNTTPLPAKVAGEGIVQLLF--LEGPRPEVTYDRKRGKYQG 172

RESULT 15

DCD_SULTO DCD_SULTO STANDARD; PRT; 183 AA.
AC Q976G3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable deoxycytidine triphosphate deaminase (EC 3.5.4.13) (dCTP deaminase).
DE deaminase).
GN Name=dcd; OrderedLocusNames=ST0226;
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Kashiwa M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
CC -!- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -!- PATHWAY: De novo synthesis of thymidylate. family.
CC -!- SIMILARITY: Belongs to the dCTP deaminase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: AP000981; BAB65184.1; ALT_INIT.
DR HSSP: Q57872; LOGH.
DR HAMAP: MF_00146; -. 1.
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR008180; DeoxyUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Complete proteome; Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 183 AA; 21055 MW; B3081C863DD71A47 CRC64;

Query Match 28.9%; Score 233; DB 1; Length 183;
Best Local Similarity 37.7%; Pred. No. 1.4e-13;
Matches 58; Conservative 27; Mismatches 55; Indels 14; Gaps 4;

QY 12 ILIEPFSESLQAGYDLRVGRE--AFVK-GKLIID-----VEKEGKVIPPPEYA 58
Db 17 IKIDPLREDTVRENGVDLAVGGEIARFIKTKVFDPNPDPAFFKIEEPIIQYEHV 76
QY 59 LILTLERIKLPDVMGDMKIRSLAREGVIGSFAMVDPQWDGNLTMLYNASNEPVELRY 118
Db 77 LITTEYIELPDNVMAFNLRSSFARLGLFIPTTIVDAGFKGQITIEVV-GSSFPVLRR 135
QY 119 GERFVQIAFIRLEGPARNPGYNGSSTRLAFSK 152
Db 136 GTRFHILIFARTLSPEVHPYQKYQGKGKVTLPK 169

Search completed: July 2, 2005, 12:24:39
Job time : 32.2558 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 2, 2005, 09:02:52 ; Search time 163 Seconds
(without alignments)
1566.007 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPWKIRKEILIEPFSEE.....PYRGNVQSTRLAFSKRKL 156

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame+ p2n.model -DB=US08957709/runat_01072005_154632_26031/app query.fasta_1.327
-O=/cgn2_1/USPTO spoal/US08957709/runat_01072005_154632_26031/app query.fasta_1.327
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human0.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US08957709 @CEN 1 1 177 @runat 01072005_154632_26031 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfileseq1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	806	100.0	471	4	US-09-399-003-70 Sequence 70, Appl
2	543	67.4	740	3	US-08-822-774-42 Sequence 42, Appl
3	543	67.4	740	3	US-09-632-711-42 Sequence 42, Appl
4	543	67.4	740	3	US-09-632-703B-42 Sequence 42, Appl
5	543	67.4	740	3	US-09-632-702-42 Sequence 42, Appl
6	543	67.4	740	4	US-09-399-003-42 Sequence 42, Appl
7	194	24.1	4403765	3	US-09-103-840A-2 Sequence 2, Appli
8	194	24.1	4411529	3	US-09-103-840A-1 Sequence 1, Appli
9	185.5	23.0	594	4	US-09-328-352-539 Sequence 539, App
10	174	21.6	640881	4	US-09-790-988-1 Sequence 1, Appli
11	172	21.3	609	4	US-09-252-991A-1091 Sequence 1091, Ap
12	168.5	20.9	579	4	US-09-540-236-1761 Sequence 1761, Ap

13	168.5	20.9	615	4	US-09-543-681A-891 Sequence 891, App
14	168.5	20.9	8947	4	US-09-596-002-34 Sequence 34, Appl
15	164	20.3	654	4	US-09-489-039A-1128 Sequence 1128, Ap
16	160	19.9	129	3	US-08-822-774-40 Sequence 40, Appl
17	160	19.9	129	3	US-09-632-711-40 Sequence 40, Appl
18	160	19.9	129	3	US-09-632-703B-40 Sequence 40, Appl
19	160	19.9	129	3	US-09-632-702-40 Sequence 40, Appl
20	160	19.9	129	4	US-09-399-003-40 Sequence 40, Appl
21	154	19.1	1230025	4	US-09-198-452A-1 Sequence 1, Appli
22	154	19.1	1230230	4	US-09-438-185A-1 Sequence 1, Appli
23	154	19.1	1664976	4	US-08-916-421B-1 Sequence 1, Appli
24	154	19.1	1664976	4	US-09-692-570-1 Sequence 1, Appli
25	149.5	18.5	1830121	4	US-09-557-884-1 Sequence 1, Appli
26	149.5	18.5	1830121	4	US-09-643-990A-1 Sequence 1, Appli
27	134.5	16.7	534	3	US-09-199-637A-246 Sequence 246, App
28	134.5	16.7	42235	3	US-09-199-637A-1 Sequence 1, Appli
29	132.5	16.4	501	3	US-09-199-637A-244 Sequence 244, App
30	123	15.3	84	3	US-08-822-774-36 Sequence 36, Appl
31	123	15.3	84	3	US-09-632-711-36 Sequence 36, Appl
32	123	15.3	84	3	US-09-632-703B-36 Sequence 36, Appl
33	123	15.3	84	3	US-09-632-702-36 Sequence 36, Appl
34	123	15.3	84	4	US-09-399-003-36 Sequence 36, Appl
35	121	15.0	444	4	US-09-583-110-1776 Sequence 1776, Ap
36	121	15.0	477	4	US-09-107-433-924 Sequence 924, App
37	118	14.6	456	4	US-09-902-540-3295 Sequence 3295, Ap
38	118	14.6	11199	4	US-09-902-540-1017 Sequence 1017, Ap
39	117.5	14.6	1664976	4	US-08-916-421B-1 Sequence 1, Appli
40	117.5	14.6	1664976	4	US-09-692-570-1 Sequence 1, Appli
41	117	14.5	10223	3	US-08-961-527-73 Sequence 73, Appl
42	112	13.9	546	3	US-09-199-637A-238 Sequence 238, App
43	111	13.8	960	2	US-08-824-405-3 Sequence 3, Appli
44	111	13.8	1037	2	US-08-824-405-1 Sequence 1, Appli
45	111	13.8	1326	4	US-09-949-016-5063 Sequence 5063, Ap

ALIGNMENTS

RESULT 1
US-09-399-003-70
; Sequence 70, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Högrefe, Holly
; APPLICANT: Hansen, Connie J
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complex
; TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying
; FILE REFERENCE: 4121.0116-02
; CURRENT APPLICATION NUMBER: US/09/399,003
; CURRENT FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
US-09-399-003-70

Alignment Scores:	1.66e-108	Length:	471
Pred. No.:	806.00	Matches:	156
Score:	100.00%	Conservatives:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	100.00%	Gaps:	0
DB:	4		

US-08-957-709A-71 (1-156) x US-09-399-003-70 (1-471)

QY 1 MetLeuLeuProAspTyrLysIleArgLysGluIleLeuIleGluProPheSerGluGlu 20
DB 1 ATGCTACTTCCAGACTGGAATATCAGAAAGAAATACCTTATAGAGCCATTTCTGAGAA 60
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGCAGAGAGGCTTTTGTAAAGGGGAAA 120
QY 41 LeuIleAspValGluLysGluLysValValIleProProArgGluTyrAlaLeuIle 60
DB 121 TTAATCGACGTGGAAGAGGAAGTTCGTTATTCCTCCAAAGGAATACGCCCTTAATC 180
QY 61 LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleArgSer 80
DB 181 CTAACCTTCGAGAGGATTAAGTTGCCGACGATGTTATGGGGATATGAAGTAAAGGAGC 240
QY 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly 100
DB 241 AGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGGCTTGGGTTGACCCAGGATGGATGA 300
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 301 AACTTAACACTAATGCTCTACAAATGCCCTCAATGAACCTGTGCAATTAAGATATGGAGAG 360
QY 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
DB 361 AGATTTGTGCAGATCGCATTTATAGGCTAGAGGGTCCGGCAGAGAAACCTTACAGAGA 420
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLeu 156
DB 421 AACTATCAGGGGAGCACAGGTTAGCGTTTCCAAAGAGAAAGAACTC 468

RESULT 2

US-08-822-774-42
; Sequence 42, Application US/08822774
; Patent No. 6183997

GENERAL INFORMATION:

; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSES: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997

CLASSIFICATION:

; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844

INFORMATION FOR SEQ ID NO:

; 42:

SEQUENCE CHARACTERISTICS:

; LENGTH: 740 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO

; ANTI-SENSE: NO
US-08-822-774-42

Alignment Scores:

Score: 1.23e-69 Length: 740
543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 3 Gaps: 0

US-08-957-709A-71 (1-156) x US-08-822-774-42 (1-740)

QY 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspVal 72
DB 2 CCTCAAGGGAATACGCCCTTAATCTTAACCTCGAGAGGATAAAAGTTGCCCGACGATGTT 61
QY 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
DB 62 ATGGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAAGGGGTTATTGGTTCTTTTGGT 121
QY 93 TrpValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
DB 122 TGGGTGTACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTACAATGCCCTCAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
DB 182 CTGTGCAGATTAAGATATGGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGCT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
DB 242 CCGGCAAGAAACCTTTACAGAGGAACTATCAGGGGAGCACAAAGGTTAGCGTTTTCAAAG 301
QY 153 ArgLysLysLeu 156
DB 302 AGAAAGAAACTC 313

RESULT 3

US-09-632-711-42
; Sequence 42, Application US/09632711
; Patent No. 6333165

GENERAL INFORMATION:

; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSES: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,711
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/822,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844

INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-632-711-42

Alignment Scores:
Pred. No.: 1,23e-69 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 3 Gaps: 0

US-08-957-709A-71 (1-156) x US-09-632-711-42 (1-740)

QY 53 ProProArgGluTyrAlaLeuLeuThrLeuGluArgLysLeuProAspVal 72
DB 2 CTCCTCAAGGGAATACGCTTATCTTACCTCGAGAGGATAAAGTTGCCGACGATGT 61
QY 73 MetGlyAspMetLysIleArgSerLeuAlaArgGluGlyValIleGlySerPheAla 92
DB 62 ATGGGGGATATGAAGATAAGGACGATTTAGCAAGAGAGGGGTATTGCTTTTGTCT 121
QY 93 TrpValAspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
DB 122 TGGGTTGACCCAGGATGGATGGAATCTTAACTAATCTTACCAATGCTTCAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGly 132
DB 182 CTGTGCAATTAGATATGAGAGAGATTTGTCAGATCGCATTTATAGGCTAGAGGT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
DB 242 CCGGCAAGAAACCTTTACAGAGGAACTATCAGGGGAGCACAAAGTTAGCGTTTCAAAG 301
QY 153 ArgLysLysLeu 156
DB 302 AGAAGAAACTC 313

RESULT 4

US-09-632-703B-42
Sequence 42, Application US/09632703B
Patent No. 6379553
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,703B
FILING DATE: 24-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/822,774

FILING DATE: 21-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: BARKER, M. Paul
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 4121.0116-04
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 849-6613
TELEFAX: (650) 849-6686
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-632-703B-42

Alignment Scores:
Pred. No.: 1,23e-69 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 3 Gaps: 0

US-08-957-709A-71 (1-156) x US-09-632-703B-42 (1-740)

QY 53 ProProArgGluTyrAlaLeuLeuThrLeuGluArgLysLeuProAspVal 72
DB 2 CTCCTCAAGGGAATACGCTTATCTTACCTCGAGAGGATAAAGTTGCCGACGATGT 61
QY 73 MetGlyAspMetLysIleArgSerLeuAlaArgGluGlyValIleGlySerPheAla 92
DB 62 ATGGGGGATATGAAGATAAGGACGATTTAGCAAGAGAGGGGTATTGCTTTTGTCT 121
QY 93 TrpValAspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
DB 122 TGGGTTGACCCAGGATGGATGGAATCTTAACTAATCTTACCAATGCTTCAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGly 132
DB 182 CTGTGCAATTAGATATGAGAGAGATTTGTCAGATCGCATTTATAGGCTAGAGGT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
DB 242 CCGGCAAGAAACCTTTACAGAGGAACTATCAGGGGAGCACAAAGTTAGCGTTTCAAAG 301
QY 153 ArgLysLysLeu 156
DB 302 AGAAGAAACTC 313

RESULT 5

US-09-632-702-42
Sequence 42, Application US/09632702
Patent No. 644428
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Purifying and Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,702
FILING DATE: 04-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/822,774
FILING DATE: 21-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-632-702-42

Alignment Scores:
Pred. No.: 1,236-69 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 3 Gaps: 0

US-08-957-709A-71 (1-156) x US-09-632-702-42 (1-740)

QY 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLeuLeuProAspAspVal 72
DB 2 CCTCAAGGGAATAGCCCTTAATCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGTT 61
QY 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
DB 62 ATGGGGGATATCAAGATAAGGAGCAGTTTAGCAGAGAGAGGTTATTGGTTCTTTTGGCT 121
QY 93 TrpValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
DB 122 TGGGTTGACCCAGGATGGGATGGAACTTAACACTAATGCTCTACAATGCCCTCAAAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
DB 182 CCTGTCGAATTAAGATATGGAGAGATTTGTGCAGATCGCATTTTATTAAGGCTAGAGGGT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
DB 242 CCGGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAAGGTTAGCGTTTTCAAAG 301
QY 153 ArgLysLysLeu 156
DB 302 AGAAGAAACTC 313

RESULT 6

US-09-399-003-42
Sequence 42, Application US/09399003
Patent No. 6734293
GENERAL INFORMATION:
APPLICANT: Hognrefe, Holly
APPLICANT: Hansen, Connie J
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying
FILE REFERENCE: 4121.0116-02

CURRENT APPLICATION NUMBER: US/09/399,003
CURRENT FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: PCT/ US98/05497
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: US 08/957,709
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: US 08/822,774
PRIOR FILING DATE: 1997-03-21
NUMBER OF SEQ ID NOS: 89
SOFTWARE: PatentIn version 3.1
SEQ ID NO 42
LENGTH: 740
TYPE: DNA
ORGANISM: Pyrococcus furiosus
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(740)
OTHER INFORMATION: "n" is a, t, g, or c
US-09-399-003-42

Alignment Scores:
Pred. No.: 1,236-69 Length: 740
Score: 543.00 Matches: 104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.37% Indels: 0
DB: 4 Gaps: 0

US-08-957-709A-71 (1-156) x US-09-399-003-42 (1-740)

QY 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLeuLeuProAspAspVal 72
DB 2 CCTCAAGGGAATAGCCCTTAATCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGTT 61
QY 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
DB 62 ATGGGGGATATCAAGATAAGGAGCAGTTTAGCAGAGAGAGGTTATTGGTTCTTTTGGCT 121
QY 93 TrpValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
DB 122 TGGGTTGACCCAGGATGGGATGGAACTTAACACTAATGCTCTACAATGCCCTCAAAATGAA 181
QY 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
DB 182 CCTGTCGAATTAAGATATGGAGAGATTTGTGCAGATCGCATTTTATTAAGGCTAGAGGGT 241
QY 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
DB 242 CCGGCAAGAAACCCCTTACAGAGGAACTATCAGGGGAGCACAAAGGTTAGCGTTTTCAAAG 301
QY 153 ArgLysLysLeu 156
DB 302 AGAAGAAACTC 313

RESULT 7

US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA

DB:	4	Gaps:	4
US-08-957-709A-71 (1-156) x US-09-328-352-539 (1-594)			
Qy	10	LyseGluIleuIleGluProPheSerGluSerLeuGln	----- 23
Db	67	AAACACGGCATGATGAACCTTATCCAGAGAAATCAAGTCGGTTTGATAGAAATGCGGAA	126
Qy	24	-----ProLaGlyTyrAspLeuArgValGlyArgGlu	--- 34
Db	127	AAATTGATTCTCAGGGGTCTCTAGCTATGGTTTATGACGTACGCTGTGCCCGCAATTT	186
Qy	35	-----AlaPheValLysGlyLys	40
Db	187	AAAGTTTTCCTACTAACGTACATCTCCAAATGTTGATCCAAAAATTTTCGATGAAAAAAGT	246
Qy	41	LeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyrAlaLeuIle	60
Db	247	TTTATCGAATTAGTCTGAC---GTTTGTATTATTCGGCTTAACCTATTTGCTTAGCG	303
Qy	61	LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer	80
Db	304	CGTACAAATTGAATATTTCGTATTCCACGTAATGTTTGTACTGTGTGCCCTGGTAATCA	363
Qy	81	SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly	100
Db	364	ACTTATGCACGTTGGCGTATTATTGTAATGTCACTCCTTTAGACCGAATGGGAAGG	423
Qy	101	AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu	120
Db	424	CAGGTTACTTTAGAAATTTCTAATACCACGAATGCTGCCGCGTATTTATGCGGCTGAA	483
Qy	121	ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArg	--- 139
Db	484	GGTGATGACCAATGCTATTTTCGAAAGTGTGAAGTATGCGAAACCTCTTATAAAGAT	543
Qy	140	-----GlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys	152
Db	544	CGTGGTGGTAATACCNAGGCCAACTGGGGTTACACTTCTCTAAG	588
RESULT 10			
US-09-790-988-1			
; Sequence 1, Application US/09790988			
; Patent No. 6632935			
; GENERAL INFORMATION:			
; APPLICANT: SHIGENOBU, SHUJI			
; APPLICANT: WATANABE, HIDEMI			
; APPLICANT: HATTORI, MASAHIRA			
; APPLICANT: SAKAKI, YOSHIYUKI			
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS			
; FILE REFERENCE: 081356/0159			
; CURRENT APPLICATION NUMBER: US/09/790,988			
; CURRENT FILING DATE: 2001-02-23			
; PRIOR APPLICATION NUMBER: JP2000-107160			
; PRIOR FILING DATE: 2000-04-07			
; NUMBER OF SEQ ID NOS: 7			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 1			
; LENGTH: 640681			
; TYPE: DNA			
; ORGANISM: Buchnera sp.			
US-09-790-988-1			
Alignment Scores:			
Pred. No.: 6,568-11 Length: 640681			
Score: 174.00 Matches: 48			
Percent Similarity: 47.56% Conservative: 30			
Best Local Similarity: 29.27% Mismatches: 58			
Query Match: 21.59% Indels: 28			
DB: 4 Gaps: 5			
US-08-957-709A-71 (1-156) x US-09-790-988-1 (1-640681)			

Db 460 GGCTGGTCCGGCTGCATCGTGTGCTTGTAGTTCTACAACTCCGGCAAGTTGCCGCTGGCGCTG 519
 Qy 117 ArgTyr-GlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
 Db 520 CGCCCGGGGATGCCCATCGGCGCGCTGAGCTTTGAACCGTTATCGGGCCCGGCGGCCCGG 579
 Qy 137 ProTyr 138
 Db 580 CCGTAT 585

Search completed: July 2, 2005, 11:27:04
 Job time : 1235 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:34:09 ; Search time 8.65116 Seconds
(without alignments)
1735.006 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 156
Sequence: 1 MLLPDWKIRKEILIEPFSE.....PYRGNVQGSTRLAFSKRKKL 156

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	23.7	156	2 E71216	dCTP deaminase (EC
2	22	14.1	154	2 G75030	dCTP deaminase (EC
3	8	5.1	420	2 D84965	dihydrolipoamide S
4	8	5.1	503	2 AB2938	alpha-L-arabinofur
5	8	5.1	503	2 E98344	hypothetical prote
6	7	4.5	133	1 I40398	flagellar protein
7	7	4.5	133	2 E84101	ail protein precu
8	7	4.5	178	2 A35123	ail protein invasio
9	7	4.5	182	2 AI0353	attachment invasio
10	7	4.5	204	2 AD3411	transporter, lyse
11	7	4.5	215	2 S74033	alkyl hydroperoxid
12	7	4.5	237	2 C64491	phosphoribosylform
13	7	4.5	268	1 A69000	conserved hypothet
14	7	4.5	279	2 A64149	hypothetical prote
15	7	4.5	327	2 A70435	cysteine synthase
16	7	4.5	352	2 C81445	probable transmemb
17	7	4.5	376	2 T40673	homoserine dehydro
18	7	4.5	439	2 I39524	probable porin - A
19	7	4.5	460	1 S48489	allantoinase (EC 3
20	7	4.5	587	2 H83748	two-component sens
21	7	4.5	661	2 A69252	3-hydroxyacyl-CoA
22	7	4.5	666	2 D82386	methyl-accepting c
23	7	4.5	682	1 G48563	18 protein - fowlp
24	7	4.5	684	2 T30652	probable RNA helic
25	7	4.5	782	2 D81281	probable nucleotid
26	7	4.5	874	2 H86167	hypothetical prote
27	7	4.5	1017	2 PC4035	cell-cycle-depende
28	7	4.5	1084	2 F95141	type II restrictio
29	7	4.5	1084	2 D98009	type II site-speci

30	7	4.5	1113	2 E64215	hypothetical prote
31	7	4.5	1921	2 T13827	kinesin-73 - fruit
32	6	3.8	34	2 S77646	hypothetical prote
33	6	3.8	35	2 C70256	hypothetical prote
34	6	3.8	37	1 R5EG36	ribosomal protein
35	6	3.8	55	2 H70228	hypothetical prote
36	6	3.8	62	2 I39622	ribosomal protein
37	6	3.8	62	2 AD1899	30S ribosomal prot
38	6	3.8	62	2 H84083	hypothetical prote
39	6	3.8	64	2 T29319	hypothetical prote
40	6	3.8	68	2 G82600	hypothetical prote
41	6	3.8	80	2 A90767	probable terminase
42	6	3.8	89	2 AI1033	hypothetical prote
43	6	3.8	94	1 VUWTEM	embryonic abundant
44	6	3.8	94	2 C97048	hypothetical prote
45	6	3.8	96	2 B82436	conserved hypothet

ALIGNMENTS

RESULT 1

E71216 dCTP deaminase (EC 3.5.4.13) PH1997 [similarity] - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C;Accession: E71216
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguch
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: E71216
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-156 <RAW>
A;Cross-references: UNIPROT:O57706; GB:AP000007; NID:G3236134; PIDN:BAA31124.1; PID:G32
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1997
C;Keywords: hydrolase

Query Match 23.7%; Score 37; DB 2; Length 156;

Best Local Similarity 100.0%; Pred. No. 1.3e-30;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLPDWKIRKEILIEPFSESLQPGAYDLRVGREAFTV 37

Db 1 MLLPDWKIRKEILIEPFSESLQPGAYDLRVGREAFTV 37

RESULT 2

G75030 dCTP deaminase (EC 3.5.4.13) PAB1164 [similarity] - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: G75030
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A;Reference number: A75001
A;Accession: G75030
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-154 <RAW>
A;Cross-references: UNIPROT:Q9UXS8; GB:AJ248288; GB:AL096836; NID:G5458960; PIDN:CAB506
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: dcd; PAB1164
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

```

Query Match      14.1%; Score 22; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 4.9e-15; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0;

QY      89 GSFANVDPGWDGNLTMLYNAS 110
      |||||
DB      89 GSFANVDPGWDGNLTMLYNAS 110

RESULT 3
DB4965
dihydrolipoamide S-succinyltransferase (EC 2.3.1.61) [imported] - Buchnera sp. (strain A
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: DB4965
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A:Reference number: AB4930; MUID:20445173; PMID:10993077
A:Accession: DB4965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <STO>
A:Cross-references: GB:AP000398; GSPDB:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: sucB; BU303
C:Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology
C:Keywords: acyltransferase; coenzyme A

Query Match      5.1%; Score 8; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 3.9; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY      45 EKEGKVI 52
      |||||
DB      62 EKEGKVI 69

RESULT 4
AB2938
alpha-L-arabinofuranosidase Atu3104 [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AB2938
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2938
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-503 <KUR>
A:Cross-references: UNIPROT:Q8UBB2; GB:AE008689; PIDN:AAL43920.1; PID:gl7741471; GSPDB:C
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3104
A:Map position: linear chromosome

Query Match      5.1%; Score 8; DB 2; Length 503;
Best Local Similarity 100.0%; Pred. No. 4.6; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY      25 AGYDLRVG 32
      |||||
DB      455 AGYDLRVG 462

RESULT 5

```

```

E98344
hypothetical protein AGR_L_3408 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: E98344
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: E98344
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-503 <KUR>
A:Cross-references: UNIPROT:Q8UBB2; GB:AE007870; PIDN:AAK90279.1; PID:gl15160302; GSPDB:C
C:Genetics:
A:Gene: AGR_L_3408
A:Map position: linear chromosome

Query Match      5.1%; Score 8; DB 2; Length 503;
Best Local Similarity 100.0%; Pred. No. 4.6; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY      25 AGYDLRVG 32
      |||||
DB      455 AGYDLRVG 462

RESULT 6
I40398
flagellar protein flis - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: I40398; E69625
R:Chen, L.; Helmann, J.D.
J. Bacteriol. 176, 3093-3101, 1994
A:Title: The Bacillus subtilis sigma D-dependent operon encoding the flagellar proteins
A:Reference number: I40396; MUID:94252974; PMID:8195064
A:Accession: I40398
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-133 <RES>
A:Cross-references: UNIPROT:P39739; EMBL:Z31376; NID:g499379; PIDN:CAA83249.1; PID:g4993
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Frit, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidos, A.; Lardinola,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maguda, S.; Maueel
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell
Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleith, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serio
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: AG9580; MUID:98044033; PMID:9384377
A:Accession: E69625
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-133 <KUN>
A:Cross-references: GB:Z99122; GB:AL009126; NID:g2636029; PIDN:CAB15550.1; PID:g2636059
A:Experimental source: strain 168
C:Genetics:
A:Gene: flis
C:Superfamily: flagellar protein flis

Query Match      4.5%; Score 7; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 16; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

```

QY 102 LTLMLYN 108
Db 23 LTLMLYN 29

RESULT 7

E84101
flagellar protein flis [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: E84101
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: E84101
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-133 <STO>
A:Cross-references: UNIPROT:Q9K6W1; GB:AP001519; GB:BA000004; NID:gl0176109; PIDN:BA073
A:Experimental source: strain C-125
C:Genetics:
A:Gene: flis
C:Superfamily: flagellar protein flis

Query Match 4.5%; Score 7; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LTLMLYN 108
Db 23 LTLMLYN 29

RESULT 8 *

A35123
all protein precursor - Yersinia enterocolitica
C:Species: Yersinia enterocolitica
C:Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 09-Jul-2004
C:Accession: A35123
R:Miller, V.L.; Bliaska, J.B.; Falkow, S.
J. Bacteriol. 172, 1062-1069, 1990
A:Title: Nucleotide sequence of the Yersinia enterocolitica Ail gene and characterizatio
A:Reference number: A35123; MUID:90130261; PMID:1688838
A:Accession: A35123
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-178 <MIL>
A:Cross-references: UNIPROT:P16454; GB:M29945; NID:gl55437; PIDN:AAA88694.1; PID:gl55438
C:Superfamily: phage lambda membrane protein lom

Query Match 4.5%; Score 7; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 GVIGSFA 92
Db 65 GVIGSFA 71

RESULT 9

A10353
attachment invasion locus protein [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: A10353
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.
Geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: A10353

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-182 <KUR>
A:Cross-references: UNIPROT:Q8ZCR3; GB:AL590842; PIDN:CAC92156.1; PID:gl5980871; GSPDB:
C:Genetics:
A:Gene: ail
C:Superfamily: phage lambda membrane protein lom

Query Match 4.5%; Score 7; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 GVIGSFA 92
Db 68 GVIGSFA 74

RESULT 10

AD3411
transporter, lyseE family BME11274 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AD3411
R:DelVecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella meliten
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3411
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <KUR>
A:Cross-references: UNIPROT:Q8YGB7; GB:AE008917; PIDN:AAL52455.1; PID:gl7983261; GSPDB:
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME11274
A:Map position: 1
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)

Query Match 4.5%; Score 7; DB 2; Length 204;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RLEGPAR 135
Db 138 RLEGPAR 144

RESULT 11 *

S74033
alkyl hydroperoxide reductase homolog - Sulfolobus solfataricus
N:Alternate names: protein c0215
C:Species: Sulfolobus solfataricus
C:Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 09-Jul-2004
C:Accession: S74033
R:Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.
Mol. Microbiol. 22, 175-191, 1996
A:Title: Organizational characteristics and information content of an archaeal genome:
A:Reference number: S73076; MUID:97055432; PMID:8899719
A:Accession: S74033
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-215 <SEN>
A:Cross-references: UNIPROT:P95895; EMBL:Y08256; NID:gl707679; PIDN:CAA69447.1; PID:gl7
A:Experimental source: strain P2
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1996
C:Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase C22 pro

Query Match 4.5%; Score 7; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 RIKLPDD 71

Db 24 RIKLPPD 30
|||||
RESULT 12
C64491
phosphoribosylformimino-5-aminomidazole carboxamide ribotide isomerase - Methanococcus
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: C64491
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.;
Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
raon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: C64491
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-237 <BUL>
A:Cross-references: UNIPROT:Q58927; GB:U67594; GB:L77117; NID:gl592160; PIDN:AAB99553.1;
C:Genetics:
A:Map position: FOR1508884-1509597
C:Superfamily: N-(5'-phospho-D-riboseylformimino)-5-amino-1-(5'-phosphoribosyl)-4-imidaz
Query Match 4.5%; Score 7; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;
Indels 0;
Qy 46 KEGKVI 52
|||||
Db 131 KEGKVI 137
RESULT 13
A69000
conserved hypothetical protein MTH1 - Methanobacterium thermoautotrophicum (strain Delta
C:Species: Methanobacterium thermoautotrophicum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A69000
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.
Xt, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: A69000
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-268 <MTH>
A:Cross-references: UNIPROT:Q26109; GB:AE000795; GB:AE000666; NID:g2621036; PIDN:AAB8451
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1
C:Superfamily: conserved hypothetical protein MTH1
Query Match 4.5%; Score 7; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;
Indels 0;
Qy 10 KEILIEP 16
|||||
Db 149 KEILIEP 155
RESULT 14
A64149
hypothetical protein HI0345 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
C:Accession: A64149
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, R.

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: A64149
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-279 <TGR>
A:Cross-references: UNIPROT:P44652; GB:U32719; GB:L42023; NID:gl573310; PIDN:AAC22006.1;
A:Note: best homolog was a hypothetical protein from Escherichia coli
C:Superfamily: conserved hypothetical protein HI0345; ferredoxin 2[4Fe-4S] homology
P:65-130/Domain: ferredoxin 2[4Fe-4S] homology <PER>

Query Match 4.5%; Score 7; DB 2; Length 279;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;
Indels 0;

Qy 81 SLAREGV 87
|||||
Db 48 SLAREGV 54
|||||

RESULT 15
A70435
cysteine synthase (EC 4.2.99.8) cysM [similarity] - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: A70435
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: A70435
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-327 <AQF>
A:Cross-references: UNIPROT:O67507; GB:AE000745; NID:g2983907; PIDN:AAC07459.1; PID:g298
A:Experimental source: strain VF5
C:Genetics:
A:Gene: cysM
C:Superfamily: threonine dehydratase
C:Keywords: carbon-oxygen lyase; phosphoprotein; pyridoxal phosphate
F:65/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted

Query Match 4.5%; Score 7; DB 2; Length 327;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;
Indels 0;

Qy 46 KEGKVI 52
|||||
Db 83 KEGKVI 89
|||||

Search completed: July 2, 2005, 12:38:51
Job time : 11.6512 secs

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	102	25.3	402	2	Q8X250	Q8X250 pyrococcus	
2	72	17.9	401	2	O59114	O59114 pyrococcus	
3	60	14.9	401	2	Q9V0S3	Q9V0S3 pyrococcus	
4	12	3.0	394	2	Q9X213	Q9X213 thermotoga	
5	12	3.0	410	2	Q83NK7	Q83nk7 tropheryma	
6	12	3.0	417	2	Q83MW7	Q83mw7 tropheryma	
7	11	2.7	403	1	DFP METJA	Q58323 m coenzyme	
8	10	2.5	86	2	Q971S3	Q971s3 sulfolobus	
9	10	2.5	394	2	Q9UT98	Q91t98 neisseria m	
10	10	2.5	394	2	Q9UTB7	Q91yb7 neisseria m	
11	10	2.5	399	2	Q9RQH7	Q9rqh7 listeria mo	
12	10	2.5	399	2	Q8Y674	Q8y674 listeria mo	
13	10	2.5	399	2	Q92AI3	Q92ai3 listeria in	
14	10	2.5	399	2	Q71VJ1	Q71vj1 listeria mo	
15	10	2.5	412	2	Q976C4	Q976c4 sulfolobus	
16	9	2.2	382	2	Q9RH70	Q9rh70 methanobact	
17	9	2.2	404	2	O28628	O28628 archaeoglob	
18	9	2.2	409	2	Q82UM0	Q82um0 nitrosomona	
19	9	2.2	414	2	O6NH19	Q6nh19 corynebacte	
20	9	2.2	431	2	Q8TRA0	Q8tra0 methanosarc	
21	9	2.2	1185	2	Q94H40	Q94h40 oryza sativ	
22	8	2.0	118	2	Q720W6	Q720w6 listeria mo	
23	8	2.0	127	2	Q9XUT2	Q9xut2 caenorhabdi	
24	8	2.0	142	2	Q9RIA1	Q9ria1 mus musculu	
25	8	2.0	169	2	Q9ZSW7	Q9zsw7 hamamelis v	
26	8	2.0	178	2	Q72JB0	Q72jb0 thermus the	
27	8	2.0	206	2	Q7Q629	Q7q629 anopheles g	
28	8	2.0	210	2	Q91596	Q91596 xenopus lae	
29	8	2.0	210	2	O6GNQ5	Q6gnq5 xenopus lae	
30	8	2.0	230	2	O6LUU4	O61uu4 photobacter	
31	8	2.0	227	1	GPWA_HAEDU	Q7V128 haemophilus	

```

DE Hypothetical protein PH1444.
GN OrderedLocusName=PH1444;
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3; PubMed=9679194;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Maeuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000006; BAA30551.1; -.
DR PIR; G71018; G71018.
DR HSP; F30197; IG63.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DRP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
SQ Complete proteome; Hypothetical protein.
KW SEQUENCE 401 AA; 44624 MW; 26CBC5F523AE02C8 CRC64;

Query Match 17.9%; Score 72; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 3.4e-61; Mismatches 0; Indels 0; Gaps 0;
Matches 72; Conservative 0;

QY 68 TGNPVITEITGRIEVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTTAPP 127
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
69 TGNPVITEITGRIEVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTTAPP 128

QY 128 HIPIMIAPAMHE 139
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
129 HIPIMIAPAMHE 140

RESULT 3
ID Q9V0S3 PRELIMINARY; PRT; 401 AA.
AC Q9V0S3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Dfp DNA/pantothenate metabolism flavoprotein.
GN ORFNames=PAB1897;
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RX MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weisenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
DR EMBL; AJ248285; CAB49630.1; -.
DR PIR; E75114; E75114.
DR HSP; P30197; IG63.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.

```

```

DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 401 AA; 44290 MW; 96CCD5A190C1FA5D CRC64;

Query Match 14.9%; Score 60; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 1.8e-49;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 IEHVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTTAFPHIPIAMIAPAMHE 139
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
81 IEHVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTTAFPHIPIAMIAPAMHE 140

RESULT 4
ID Q9X213 PRELIMINARY; PRT; 394 AA.
AC Q9X213;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE DNA/pantothenate metabolism flavoprotein.
GN OrderedLocusName=TW1687;
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=99287316; PubMed=10360571; DOI=10.1038/20601;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L.A., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.L.,
RA Heidelberg J.F., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001809; AAD36754.1; -.
DR PIR; A72223; A72223.
DR HSP; Q9SWES; IE20.
DR TIGR; TM1687; -.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 394 AA; 43804 MW; EBCC811C151ECFCA CRC64;

Query Match 3.0%; Score 12; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 RFTTNASSGKMG 223
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
200 RFTTNASSGKMG 211

RESULT 5
ID Q83NK7 PRELIMINARY; PRT; 410 AA.
AC Q83NK7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative flavoprotein.
GN OrderedLocusName=TW408;
OS Tropheryma whipplei (strain TW08/27) (Whipple's bacillus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

```

OC Micrococcineae; Cellulomonadaceae; Tropheryma.

OX NCBI_TaxID=218496;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22495039; PubMed=12606174; DOI=10.1016/S0140-6736(03)12597-4;

RA Bentley S.D., Maiwald M., Murphy L.D., Pallen M.J., Yeats C.A.,

RA Dover L.G., Norbertcak H.T., Besra G.S., Quail M.A., Harris D.E.,

RA von Herbay A., Goble A., Rutter S., Squares R., Squares S.,

RA Barrell B.G., Parkhill J., Rellman D.A.,

RT "Sequencing and analysis of the genome of the Whipple's disease

RT bacterium Tropheryma whippelii";

RL Lancet 361:637-644 (2003).

DR EMBL; BX251411; CAD67079.1; -.

DR HSSP; QSW55; IE20.

DR InterPro; IPR005252; Cons hypoth521.

DR InterPro; IPR007085; Dfp_C.

DR InterPro; IPR003382; Flavoprotein.

DR Pfam; PF04127; Dfp; 1.

DR Pfam; PF02441; Flavoprotein; 1.

DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.

KW Complete proteome.

SQ SEQUENCE 410 AA; 44783 MW; EC7BAFB9F23DDB29 CRC64;

Query Match 3.0%; Score 12; DB 2; Length 410;

Best Local Similarity 100.0%; Pred. No. 0.013; 0; Indels 0; Gaps 0;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 DVVIMAAVSDF 288

|||||

Db 268 DVVIMAAVSDF 279

RESULT 6

Q83MW7

ID Q83MW7 PRELIMINARY; PRT; 417 AA.

AC Q83MW7

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Flavoprotein.

GN Name=dpf; OrderedLocNames=TW1361;

OS Tropheryma whippelii (strain Twist) (Whipple's bacillus).

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Micrococcineae; Cellulomonadaceae; Tropheryma.

OX NCBI_TaxID=203267;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Twist;

RX MEDLINE=22784088; PubMed=12902375;

RA Raoult D., Ogata H., Audic S., Robert C., Suhr K., Drancourt M.,

RA Claverie J.-M.,

RT "Tropheryma whippelii Twist: a human pathogenic Actinobacteria with a

RT reduced genome";

RL Genome Res. 13:1800-1809 (2003).

DR EMBL; AS016851; AAO44458.1; -.

DR HSSP; Q9RC23; 1P3Y.

DR InterPro; IPR005252; Cons hypoth521.

DR InterPro; IPR007085; Dfp_C.

DR InterPro; IPR003382; Flavoprotein.

DR Pfam; PF04127; Dfp; 1.

DR Pfam; PF02441; Flavoprotein; 1.

DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.

KW Complete proteome.

SQ SEQUENCE 417 AA; 45584 MW; E2C79EC9B860BC63 CRC64;

Query Match 3.0%; Score 12; DB 2; Length 417;

Best Local Similarity 100.0%; Pred. No. 0.013;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 DVVIMAAVSDF 288

|||||

Db 275 DVVIMAAVSDF 286

RESULT 7

Dfp_METJA

ID Dfp_METJA

AC Q58323;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Coenzyme A biosynthesis bifunctional protein coaBC (DNA/pantothenate

DE metabolism flavoprotein) [Includes: Phosphopantothenoylecysteine

DE decarboxylase (EC 4.1.1.36) (PPDCD) (CoAC); Phosphopantothenate--

DE cysteine ligase (EC 6.3.2.5) (Phosphopantothenoylecysteine synthase)

DE (PPC synthetase) (PPCS) (Coab)].

GN Name=coaBC; OrderedLocNames=MJ0913;

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;

OC Methanocaldococcaceae; Methanocaldococcus.

OX NCBI_TaxID=2190;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,

RA Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,

RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.W., Glodek A.,

RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,

RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,

RT "Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii.";

RL Science 273:1058-1073 (1996).

CC -!- FUNCTION: Catalyzes two steps in the biosynthesis of coenzyme A.

CC In the first step cysteine is conjugated to 4'-phosphopantothenate

CC to form 4'-phosphopantothenoylecysteine, in the latter compound is

CC decarboxylated to form 4'-phosphopantothene (By similarity).

CC -!- CATALYTIC ACTIVITY: N-[(R)-4'-phosphopantothenoylecysteine =

CC pantotheine 4'-phosphate + CO(2).

CC -!- CATALYTIC ACTIVITY: CTP + (R)-4'-phosphopantothenoylecysteine

CC = CMP + Ppi + N-[(R)-4'-phosphopantothenoylecysteine.

CC -!- COFACTOR: Binds 1 FMN per subunit (By similarity).

CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; second step.

CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; third step.

CC -!- SUBUNIT: Homododecamer, the coaB domains form homodimers.

CC -!- SIMILARITY: In the N-terminal section; belongs to the HFCD (homo-

CC oligomeric flavin containing Cys decarboxylase) superfamily.

CC -!- SIMILARITY: In the C-terminal section; belongs to the PPC

CC synthetase family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U67535; AAB98918.1; -.

DR PIR; A64414; A64414.

DR HSSP; P30197; IG63.

DR TIGR; MJ0913; -.

DR InterPro; IPR011256; Bac reg effector.

DR InterPro; IPR005252; CoaB hypoth521.

DR InterPro; IPR007085; Dfp_C.

DR InterPro; IPR003382; Flavoprotein.

DR Pfam; PF04127; Dfp; 1.

DR Pfam; PF02441; Flavoprotein; 1.

DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.

KW Complete proteome; Flavoprotein; FMN; Hypothetical protein; Ligase;

KW Lyase.

CHAIN 1 195 Phosphopantothenoylecysteine

decarboxylase.

FT

```

FT CHAIN      196 403      Phosphopantothenate--cysteine ligase.
FT ACT SITE   87 87      By similarity.
SQ SEQUENCE   403 AA; 45670 MW; B8851AA81A115B94 CRC64;

Query Match      2.7%; Score 11; DB 1; Length 403;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSGKMGVALAE 228
Db 219 SSGKMGVALAE 229
|||||

RESULT 8
Q971S3 ID Q971S3 PRELIMINARY; PRT; 86 AA.
AC Q971S3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein STS144.
GN ORFNames=STS144;
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AF000985; BAB66347.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 86 AA; 10227 MW; CB80DBC0FD13B93 CRC64;

Query Match      2.5%; Score 10; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 IERLKKLGVE 160
Db 49 IERLKKLGVE 58
|||||

RESULT 9
Q9J798 ID Q9J798 PRELIMINARY; PRT; 394 AA.
AC Q9J798;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein NMA1916.
GN OrderedLocNames=NMA1916;
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=2022556; PubMed=10761919; DOI=10.1038/35006655;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.M.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K.L., Quail M.A.,

```

```

RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.";
RL Nature 404:502-506(2000).
DR EMBL; AL162757; CAB85137.1; -.
DR PIR; D81819; D81819.
DR HSP; Q9SWE5; 1E20.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 394 AA; 42238 MW; E19917C24E62B087 CRC64;

Query Match      2.5%; Score 10; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSGKMGVALA 227
Db 208 SSGKMGVALA 217
|||||

RESULT 10
Q9JYB7 ID Q9JYB7 PRELIMINARY; PRT; 394 AA.
AC Q9JYB7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE DNA/pantochenate metabolism flavoprotein.
GN OrderedLocNames=NMB1658;
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307; DOI=10.1126/science.287.5459.1809;
RA Tettelin H., Saunders N.J., Heidelberg J.F., Jeffries A.C.,
RA Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F.,
RA Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R.T., Peterson J.D.,
RA Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D.,
RA Dougherty B.A., Mason T.M., Ciecko A., Parksey D.S., Blair E.,
RA Cittone H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H.M.,
RA Qin H., Vamathevan J.J., Gill J., Scarlato V., Maignani V., Pizza M.,
RA Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R.,
RA Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002516; AAF42007.1; -.
DR PIR; A81058; A81058.
DR HSP; Q9SWE5; 1E20.
DR TIGR; NMB1658; -.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 394 AA; 42212 MW; B427D1901932AB91 CRC64;

Query Match      2.5%; Score 10; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSGKMGVALA 227

```


Db 208 SSGXGVALA 217
|||||

RESULT 11
Q9RQH7 PRELIMINARY; PRT; 399 AA.

ID Q9RQH7
AC Q9RQH7 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Pantothenate metabolism flavoprotein homolog.
GN Name=dpf;
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD;
RX MEDLINE=20208566; PubMed=10746777;
RT "Identification of new loci involved in adhesion of Listeria
monocytogenes to eukaryotic cells.";
RL Microbiology 146:731-739(2000).
DR EMBL; AF104226; AAF04763.1; -
DR HSP; Q9SWES; 1E20.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
SQ SEQUENCE 399 AA; 43395 MW; 566E98FF6F07F7EB CRC64;

Query Match 2.5%; Score 10; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 VLVTAGATRE 206
Db 190 VLVTAGATRE 199
|||||

RESULT 12
Q8Y674 PRELIMINARY; PRT; 399 AA.

ID Q8Y674
AC Q8Y674 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Lm1825 protein.
GN OrderedLocusNames=lmo1825;
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669; DOI=10.1126/science.1063447;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
Etian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
Rammel B., Rose M., Schluter T., Simoes N., Tierrez A.,
Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
"Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591981; CAC99903.1; -
DR PIR; A11674; A11674.
DR HSP; Q9SWES; 1E20.
DR Lm1825 protein.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 399 AA; 43475 MW; 2FEAD8EF395981F5 CRC64;

Query Match 2.5%; Score 10; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 VLVTAGATRE 206
Db 190 VLVTAGATRE 199
|||||

RESULT 14
Q71YJ1 PRELIMINARY; PRT; 399 AA.

ID Q71YJ1
AC Q71YJ1

DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Phosphophenolpyruvate decarboxylase/phosphoenolpyruvate--
DE cytosine ligase.
GN Name=coaBC; OrderedLocusNames=LMOF2365.1853;
OS Listeria monocytogenes (serotype 4b / strain F2365).
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=263669;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15115801; DOI=10.1093/nar/gkh562;
RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,
RA Kolonay J.P., Rasko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,
RA Peterson J.D., White O., Nelson W.C., Nierman W.C., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,
RA Haft D.H., Selengut J., Van Aken S.E., Khouri H.M., Fedorova N.,
RA Forberger H.A., Tran B., Kathariou S., Wonderling L.D., Uhlich G.A.,
RA Bayles D.O., Luchansky J.B., Fraser C.M.;
RT "Whole genome comparisons of serotype 4b and 1/2a strains of the food-
borne pathogen Listeria monocytogenes reveal new insights into the
core genome components of this species."
RL Nucleic Acids Res. 32:2386-2395(2004).
DR EMBL; AE017328; AAT04623.1; --
DR GO; GO:0016874; F.ligase activity; IEA.
DR InterPro; IPR005252; Cons hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome; Ligase.
SQ SEQUENCE 399 AA; 43490 MW; 1EF88D4C8C2A8344 CRC64;

Query Match 2.5%; Score 10; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVVTAGATRE 206
Db |||||
190 VLVVTAGATRE 199

RESULT 15
Q976C4 PRELIMINARY; PRT; 412 AA.
AC Q976C4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein ST0257.
GN OrderedLocusNames=ST0257;
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JCM 10545 / 7;
RA MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermophilic
Crenarchaeon, Sulfolobus tokodaii strain 7."
RL DNA Res. 8:123-140(2001).
DR EMBL; AF000982; BAB5223.1; --
DR InterPro; IPR005252; Cons hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.

DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 412 AA; 45517 MW; E2B28B9F11338947 CRC64;

Query Match 2.5%; Score 10; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 PATANTISKI 109
Db |||||
98 PATANTISKI 107

Search completed: July 2, 2005, 12:38:11
Job time : 82.0233 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:17:13 ; Search time 23.0698 Seconds
(without alignments)
1680.787 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 1998
Sequence: 1 MLHHVKKLIYATSKRKLVGKK.....KKKKRELAERIWDIEIKXLS 403

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1667.5	83.5	401	2 G71018	pantothenate metab
2	1649.5	82.6	401	2 E75114	dna/pantothenate m
3	735	36.8	404	2 D69455	pantothenate metab
4	731	36.6	386	2 D69029	pantothenate metab
5	648.5	32.5	413	2 B90162	DNA/pantothenate m
6	627.5	31.4	403	2 A64414	pantothenate metab
7	626	31.3	392	2 C84215	pantothenate metab
8	605	30.3	398	2 C97112	flavoprotein invol
9	543.5	27.2	390	2 C70201	pantothenate metab
10	535.5	26.8	388	2 B70371	pantothenate metab
11	535	26.8	437	2 A72498	probable DNA/panto
12	506	25.3	399	2 A11674	pantothenate metab
13	494.5	24.7	404	2 F83963	flavoprotein dfp l
14	494	24.7	399	2 A11302	pantothenate metab
15	490.5	24.5	399	2 A82351	DNA/pantothenate m
16	483.5	24.2	402	2 B82982	DNA/pantothenate m
17	482.5	24.1	394	2 A72223	pantothenate metab
18	473.5	23.7	406	2 D69878	pantothenate metab
19	473	23.7	394	2 A81058	DNA/pantothenate m
20	464	23.2	394	2 D81819	hypothetical prote
21	459.5	23.0	400	2 G64104	pantothenate metab
22	457	22.9	405	2 A10006	DNA/pantothenate m
23	453.5	22.7	399	2 F89893	hypothetical prote
24	449	22.5	404	2 A13538	phosphopantothenoy
25	447	22.4	422	2 H75501	DNA/pantothenate m
26	439	22.0	402	2 S75082	pantothenate metab
27	437.5	21.9	401	2 AD2615	pantthotenate metab
28	435.5	21.8	410	2 AE2194	pantothenate metab
29	430.5	21.5	418	2 H70899	pantothenate metab

30 424.5 21.2 430 1 A65165 pantothenate metab
31 424.5 21.2 430 2 C86040 pantothenate metab
32 424.5 21.2 430 2 B91193 pantothenate metab
33 420.5 21.0 407 2 AC0971 conserved hypochet
34 418.5 20.9 419 2 G86976 probable flavoprot
35 408 20.4 367 2 C97397 pantothenate metab
36 404.5 20.2 409 2 H82842 DNA/pantothenate m
37 397 19.9 412 2 F87709 DNA/pantothenate m
38 351 17.6 425 2 E71889 pantothenate metab
39 350 17.5 425 2 A64625 pantothenate metab
40 301 15.1 384 2 F81354 DNA /pantothenate
41 260 13.0 183 2 H95142 flavoprotein limpo
42 260 13.0 183 2 F98010 N-terminal region
43 239 12.0 234 2 E98010 Dfp protein homolo
44 238 11.9 229 2 G95142 conserved hypochet
45 224.5 11.2 231 2 E86696 flavoprotein limpo

ALIGNMENTS

Result 1
G71018
pantothenate metabolism flavoprotein dfp homolog PH1444 - Pyrococcus horikoshii
N;Alternate names: probable aspartate 1-decarboxylase activase
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C;Accession: G71018
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguch
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: G71018
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-401 <RAW>
A;Cross-references: UNIPROT:O59114; GB:AP0000006; MID:g3236133; PIDN:BAA30551.1; PID:g32
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1444

Query Match 83.5%; Score 1667.5; DB 2; Length 401;
Best Local Similarity 83.4%; Pred. No. 3.4e-92;
Matches 337; Conservative 33; Mismatches 29; Indels 5; Gaps 4;

Qy 1 MLHHVKKLIYATSKRKLVGKKIVXXPGSTAALD-VKACEGLIRHGAEVHVMSEAATKII 59
Db 1 MLHHVKKRIYAKSRKLVGKKIVLAI PGSTAACVCLARELIRHGAEVHVMTPSATKII 60
Qy 60 HPYAMNLPNGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDDPVT 119
Db 61 HPYAMEFATGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDDPVT 120
Qy 120 TVVTTAPPHIPMIAPAMHETWYRHPIVRENIRLKKLGVEFTGPRIEGRKAVASIDEI 179
Db 121 TVVTTAPPHIPMIAPAMHESMYKHPIVRENIEKLLKLGVEFTGPRIEGRKAVASIDEI 180
Qy 180 VYRVIKLHKKTLEGRKVLVTAGATREYIDPFRITNASSGKMGVALAEADPRGA-VTL 238
Db 181 VYRVIRKLPKTKLGRVLVTAGATREYIDPFRITNASSGKMGVALAEAEPRGAVTL 240
Qy 239 IRTKGSVKAFRIKRLKLVETVEMLSAIENELRSKKYDVVIMAAVSDPFRPKIKAEKGI 298
Db 241 IKTKGSVNSFVENQI--QVETVEMLAAIEKELTEKYYDVVIMAAVSDPFRPKIKAEKDI 298
Qy 299 KSGRSITIELVPXNPKIIDRIEIQPNVFLVGKASTSEKLEIEGKQRIERAKADLVVG 358
Db 299 KSKNSITIELVP-NPKIIDRIKELQDPVFLVGKASTMEKLIQEAQKQIERAKSDIVIG 357
Qy 359 NILEAFSGSENVVLIGRDFTKELPKMKRELAERIWDIEIKXL 402

Db 358 NTLEAFSGDESKVIIIGRNFEXKELPKMKRELAERIWDIEIKRL 401

RESULT 2

D69455

dnal/pantothenate metabolism flavoprotein (dfp) PAB1897 - Pyrococcus abyssi (strain Orsay

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

C:Accession: E75114

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: E75114

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-401 <KAW>

A:Cross-references: UNIPROT:Q9V0S3; GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB4963

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1897

C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 82.6%; Score 1649.5; DB 2; Length 401;

Best Local Similarity 82.9%; Pred. No. 4e-91;

Matches 335; Conservative 31; Mismatches 33; Indels 5; Gaps 4;

QY 1 MLHVKLIYATSKRLVGVKKIIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEATKII 59

DB 1 MLSHKLIYATSKRLVGVKKIIVLAIPGSTAAVECVKLARELIRHGAEVHVMSPSATKII 60

QY 60 HPYANNLPTGNPVITEITGFIHVELAGEHENKADLIIVCPATANTISKIACGIDTPTV 119

DB 61 HPYAMEFATGNPVITEITGSIHVELAGEHENKADLIIVCPATANTISKIACGIDTPTV 120

QY 120 TVVTTAFPHIPIAMIAPAMHETMYRHPVIRENTERLKKLVGFVFGPRIEEGRAKVASIDEI 179

DB 121 TVVTTAFPHIPIAMIAPAMHESMYKHPVIRENTERLKKLVGFVFGPRIEEGRAKVASIDEI 180

QY 180 VYRVTKLHKHTLEGKRVLVTAGATREYIDPIRFTINASSGKMGVALAEADFRGA-VTL 238

DB 181 VYRVTKLHKHTLEGKRVLVITAGATREYIDPIRFTINASSGKMGVALAEADFRGA-VTL 240

QY 239 IRTKGSVKAFRIKIKLVETVEMLSAENELRSKKYDVVIMAAVSDFRPKIAEKGI 298

DB 241 IRTGSVKSFVENQI--EVETVEMLGAEELTKKKYDVVIMAAVSDFRPKIAEKGI 298

QY 299 KSGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKKELIEGKQIERAKADLVVG 358

DB 299 KSDKSITIELVP-NPKIIRKIEIQSDVFLVGFKAETSKKELIEGKQIESAGSDLVIG 357

QY 359 NTLEAFSGSENVQLIGRDTKELPKMKRELAERIWDIEIKL 402

DB 358 NTLEAFSGSENVIVGKDFVKELPKMKRELAERIWDIEIKI 401

RESULT 3

D69455

pantothenate metabolism flavoprotein dfp homolog AF1645 - Archaeoglobus fulgidus

N:Alternate names: probable aspartate 1-decarboxylase activase

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004

C:Accession: D69455

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.P.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Wosce, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo

A:Reference number: A69250; NID:98049343; PMID:9389475

A:Accession: D69455

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-404 <KLE>

A:Cross-references: UNIPROT:O28628; GB:AE000989; GB:AE000782; NID:g2689312; PIDN:AA88955

C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 36.8%; Score 735; DB 2; Length 404;

Best Local Similarity 45.0%; Pred. No. 1.2e-36;

Matches 182; Conservative 74; Mismatches 122; Indels 26; Gaps 13;

QY 4 HVKLIYATSKRLVGVKKIIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEATKIIHPY 62

DB 5 HLERIRGRSRKLERKKIIVLGVTSIAAVETVKLARELVRRGADVAVNSRAARKIIHPY 64

QY 63 AWWLPTGNPVITEITGFIHVELAGEHENKADLIIVCPATANTISKIACGIDTPTVTV 122

DB 65 ALFPATGKRVVTEITGSIHVNLLGEY-GDADLFIACTANTISKIAGIDTPTVTPA 123

QY 123 TTAF-PHIPIMIAPAMHETMYRHPVIRENTERLKKLVGFVFGPRIEEGRAKVASIDEI 181

DB 124 TTALSGKPIIIVPAMHEAMRNKAVLENIQRLIDMGIEFVQPRIESEGKAFSTETICL 183

QY 182 RVIKKHKHTLEGKRVLVTAGATREYIDPIRFTINASSGKMGVALAEADFRGA-VTLIR 240

DB 184 HVRELYPKMKGRKRVVTSPTGTEYQIDPIRFTINSSGRMGLETALEFWRGADVHT 243

QY 241 TKGSVKAFRIKIKLVETVEMLSAENELRSKKYDVVIMAAVSDFRPKIAEKGI 300

DB 244 SKPSGMSLPNYK-EIRVMSVEDMKAVLEI-GKCDLPVSSAAAADPIVDAAEK-KIKT 300

QY 301 GRSTIELVPXNPKIIDRIKEIQPNVFLVGFKAET--SKEKLIIEGKQIERAKADLVVG 358

DB 301 APELVKL-KESPKIIKEVRKIYSG-HIIGFKAETGMSDDELLKVASERKADNLMVVA 358

QY 359 NTL--EAGSSENVQLIGRDTKELPKMK-----KRELAERI 394

DB 359 NDVLERGMDTRVLIL-----TPKRQEWELGKHVAERI 395

RESULT 4

D69029

pantothenate metabolism flavoprotein dfp homolog MTH1216 - Methanobacterium thermoautot

N:Alternate names: probable aspartate 1-decarboxylase activase

C:Species: Methanobacterium thermoautotrophicum

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004

C:Accession: D69029

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;

Qi, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N

Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funci

A:Reference number: A69000; MUID:98037514; PMID:9371463

A:Accession: D69029

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-386 <MTH>

A:Cross-references: UNIPROT:O27284; GB:AE000889; GB:AE000666; NID:g2622318; PIDN:AA88571

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1216

C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 36.6%; Score 731; DB 2; Length 386;

Best Local Similarity 44.9%; Pred. No. 1.9e-36;

Matches 176; Conservative 62; Mismatches 122; Indels 32; Gaps 13;

QY 20 KIVXXXPGSIAALD-VKACEGLIRHGAEVHVMSEATKIIHPYANLFTGHPVITEIG 78

DB 2 EILCVTSVAAIEAVKLARELGQAEVKCFMSEDAERIIHPYAMEFATGSKPVLELVG 61

QY 79 EIVHVELAGEHENKADLIIVCPATANTISKIACGIDTPTVTVTAF-PHIPIMIAPAM 137

DB 62 EIEHVKYAG-----ADLIIVAPATANIIGKLAYRLADNPISLLLTASGMGTPIVMVPSM 116

```
QY 138 HETMYRHPVIRENIERLKLGVETFGPIRIEGRKAVASIDEIVYRVVKKLHKTKTGKRV 197
Db 117 HEAMY--AAAENIMLKEGVIFIEPRMDEGKAKFPDITIVLEAMRQTSQRLOGKRV 174
QY 198 LVTAGATREYIDPIRITINASSGKMGVALAEADFRGA-VTLIRTKGSV-----KAFRI 250
Db 175 LVSLGGTYEPIDPVGITNRSSGKMGLAIRRAYIEGADVTVVAGTVSVIEIPQLRSFR- 233
QY 251 RKTKLVETVEEMLSAIENELRSKKYDVIMAAVSDFRPKIKAEKGKIKSGRSITIELVP 310
Db 234 -----AETAEEAERVRELVA--HDVFISAAVADFKP-VYTERKISSSEBSFVELRP 284
QY 311 XNPKIIDRIKEIOPNVFLVGFKA--TSKEKLEEGKQRIERAKADLVVGN--TLAFCG 366
Db 285 -NPKVIGIAREINPEAFIVGFKAEDVDNEALVESARKQIRESGVDMVANDVSVGFGS 343
QY 367 EENQVVLIGDRFKELPKMKRELAERIMDEI 398
Db 344 DRNRALIVS-DMVTEPLMEKEELASIIIDEV 374

RESULT 5
B90162
DNA/pantothenate metabolism flavoprotein (dfp) [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: B90162
R;Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, B.
arrett, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: B90162
A:Molecule type: DNA
A:Residues: 1-413 <KUR>
A:Cross-references: UNIPROT:Q98050; GB:AE006641; NID:g13813345; PIDN:AAK40553.1; GSPDB:G
A:Gene: dfp
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 32.5%; Score 648.5; DB 2; Length 413;
Best Local Similarity 39.7%; Pred. No. 1.7e-31;
Matches 166; Conservative 81; Mismatches 144; Indels 27; Gaps 11;

QY 1 MLHHVKLIYATSKRLVGVKIVXXPGSTA--ALDVKACEGLIRHGAHVHVMSEAATK 57
Db 5 MTHPSKKIIGSISNELADKKILLAVTGSVAIYKSLDL--ARSLMRNGAEVSVIISKDAK 62

QY 58 IHPYAWNLPNGPVITEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDTDP 117
Db 63 LISPENFKWATGNVVTKLTDLEHVSLEADN---DVMIVAPSTANTWVKIAYGIADIP 118

QY 118 VTTVTTAFPHI-----PIIAPAMHETMYRHPVIRENIERLKLGVETFGPIRIEGRKAV 173
Db 119 IT---ATALNFVGKKPLIIVPSMHLQWVISPQVADAADRLKRIQVEVIEPEIYVGLAHY 175

QY 174 ASIDEIVYVVIK-KLHKTKLEGRKVLVTAGATREYIDPIRITINASSGKMGVALAEAD 232
Db 176 PKLEYITSRITSVYLVKGLDGFNIATAGPTREYIDSVRFINPSSGTMGISIANEAYF 235

QY 233 RGA-VTLIRTKGSVKAFRIRKIKLVETVEEMLSAIENELRSKKYDVIMAAVSDFRPK 291
Db 236 RGAQVRVICGPISSKLEPYVDIVYVETTEEMLNEVSVKSIENGKFNVLILAGAPADYKFK 295

QY 292 IKAEGKIKSGRSTTIELVPXNPKIIDRIKEIQNVFLVGFKAET--SKEKLIBEGKQRIE 349
Db 296 NKSDTKIDSHTEIPKVELERTPKISEYIRKY--NVLVGSFASVNSDEELIEKAKIKMR 353

QY 350 RAKADLVVGNLLE----ARGSENOVVLIGRD-FTKELPKMKRELAERIWDEIKXL 402
Db 354 RHGFDLIVANNRRKIDGFSSEYNEVIVIDKNGDVRKIPKNFKTVIARKILDIVKSQL 411
```

RESULT 6
A64414

pantothenate metabolism flavoprotein MJ0913 - Methanococcus jannaschii
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: A64414
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A.; Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:868087
A:Accession: A64414
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-403 <BUL>
A:Cross-references: UNIPROT:Q58323; GB:U67535; GB:L77117; NID:g2826348; PIDN:AAB989918.1
C:Genetics:
A:Map position: REV845792-844581
C:Superfamily: pantothenate metabolism flavoprotein dfp
C:Keywords: flavoprotein

Query Match 31.4%; Score 627.5; DB 2; Length 403;
Best Local Similarity 39.6%; Pred. No. 3e-30;
Matches 163; Conservative 82; Mismatches 138; Indels 29; Gaps 16;

```
QY 1 MLHHVKLIYATSKRLVGVKIVXXPGSTAALDV-KACEGLIRHGAHVHVMSEAATKII 59
Db 5 IMHPTKLLGTGTSKLLLENKKILVAVTSSIAAIEPTPLMRLEIRHGAEVYCIITEETKII 64
QY 60 HPYAWNLPNGPVITEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVPT 119
Db 65 GKELAFGCGNEVEIEITGDIIEHIL-----YNECDCLLIYPATANIISKINLGIADNVN 120

QY 120 TVVTTAFPHIPIIAPAMHETMYRHPVIRENIERLKLGVETFGPIRIEGRKAVASIDE 178
Db 121 TTALMFFGNKPIFIVPAMHENMFN--AIKRHIDKLEKDKIYIISPKEFGKAKVANIED 178

QY 179 IYRVVTKL-HKKTLEGRKVLVTAGATREYIDPIRITINASSGKMGVALAEADFRG-AV 236
Db 179 VYKAVTEKIGNLKKEGNRVLINGGTVEFIDKRVISNLSGKMGVALAEAFCKEGFYV 238

QY 237 TLIRTKGSVKAFRIRKIKLVETVEEMLS-AIENELRSKKYDVIMAAVSDFRPKIAE 295
Db 239 EVITAMGLEPPYIK--NHKVLTAKEMLNKAIE---LAKDFDIIISAAISDFTVE-SFE 292

QY 296 GKIKSGRSITIELVPXNPKIIDRIKEIQNVFLVGFKA--TSKEKLIBEGKQRIERAKA 353
Db 293 GKLSSEEEILLKL-KENPKVLEELRIYKDKVIIGFKAENVLDEKELINRAKERLNKYNL 351

QY 354 DLVVGNTLEA--FGSENOVVLIGROFTK-ELPQM--KRELAERIWDEIEK 400
Db 352 NMIIANDLSKHVGGDDYIEVYII---TKYEVEKISGSKKEISERIVEKVK 399
```

RESULT 7
C84215

pantothenate metabolism flavoprotein [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: C84215
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky,
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabb,
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L
A.; Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: C84215

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-392 <STO>
A:Cross-references: UNIPROT:Q9HRS1; GB:AE004437; NID:g10580167; PIDN:AAG19087.1; GSPDB:C
C:Genetics: dff
A:Gene: dff
C:Superfamily: pantothenate metabolism flavoprotein dff

Query Match 31.3%; Score 626; DB 2; Length 392;
Best Local Similarity 38.1%; Pred. No. 3.5e-30;
Matches 151; Conservative 73; Mismatches 146; Indels 26; Gaps 11;

QY 16 LVGKIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAATKIHPYAWNLPTGNPVIT 74
DB 2 LSGVNVAVGVTSIAAKVVEFVHELRGACVRAVMTESAGIHPWAVEPATENPVVT 61

QY 75 EITGFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVTVTVTTAF-PIHIPMI 133
DB 62 EITGVPHVELCG-RDGWADVFPVAPATANTVKGIAAAVDDSPVTTCTVTAAGVDPVV 120

QY 134 APAMHETMYRHPVIRENTERLKLGVFIEGRIEGRKAVASIDIVYRVIKLHKHTLE 193
DB 121 VPAMHEPMYDHGVRDAIDRVSSGVSVDPRIEEGKAKLPRESTIVHETARAAGEQPLA 180

QY 194 GKRVLVTAGATREYIDPIRFITNASSGKMGVALAEADFRGA-VTLIRTK---GSVKAFR 249
DB 181 GTHVVVTSGATSEADIPVRVLTRNAGRTGRAVAAACVYRGARVTLVHDASDGGAVPYAD 240

QY 250 IRKIKLVETVBEMLSAENELRSKKYDVVIMAAVSDFRPKIKAEKIKSGR-SITIEL 308
DB 241 VRDVSSAAEMTAATLTACADA-----DALVSAAISDYTVE-AADEKLRSRGEDVALSL 293

QY 309 VPXNPKIIDRIKEIQNPVFLVGFKAET-----SKEKLIBEGKQIERAKADLVVGNL 361
DB 294 EPTR-KLVGADVDDNPDLPIVGFKAETPADGDTAGDSDSMVAARSLLQRLGLAFVANDA 352

QY 362 EAFGSEENQVILGRDFTKELPKMKK---RELAERI 394
DB 353 GVMGNDETALFVTDVSDVSEYAGHKRGLGARIAERL 388

RESULT 8
C97112
flavoprotein involved in pantothenate metabolism, YLOI B. subtilis ortholog [imported]
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: C97112
R: Nolling, J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
; Daly, M.J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: C97112
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <KUR>
A:Cross-references: UNIPROT:Q97IC8; GB:AE001437; PIDN:AAK79686.1; PID:g15024686; GSPDB:C
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1720
C:Superfamily: pantothenate metabolism flavoprotein dff

Query Match 30.3%; Score 605; DB 2; Length 398;
Best Local Similarity 39.8%; Pred. No. 6.4e-29;
Matches 165; Conservative 76; Mismatches 124; Indels 50; Gaps 17;

QY 19 KKVXVXXPGSIA---ALDV-----KACEGLIRHGAHVHVMSEAATKIHPYAWNLPTGN 70
DB 4 KNVIVGSGVIAVYKALDVISRLKADFG-----VDVMTNSAKSFVPLSFQSLGN 56

QY 71 PVITEI-----TGFEHVELAGEHENKADLILVCPATANTISKIACGIDDPVTVTVTTA 125
DB 57 KVINDMFEPEKWEIOHISLA---KKAIDLIVVPATANIIGKVANGIADDMLSITMAA 112

QY 126 PPHIPIMIPAMHETMYRHPVIRENTERLKLGVFIEGRIEGRKAVASI-----DE 178
DB 113 YGRCEPIVPAPAMTMYKNPQVNNIKKLKDYGIYIFP--EKGRACGDVGEGLADTE 170

QY 179 IVYRVIKKL--HKKTLEGRVLTAGATREYIDPIRFITNASSGKMGVALAEADFRGA- 235
DB 171 LIYENIKSLYNKRDLAGKQVWVTAGPTIAPIDPVRFTTNHSTGKMGYAIASEARDGAE 230

QY 236 VTLIRTKGSVK-APRIKIKLVETVBEMLSAENELRSKKYDVVIMAAVSDFRPKIKA 294
DB 231 VTLISGETSLPPFGVDFI--KVTNSSEMMEKVLKFGSG--DIVIKSAVADYKAKNYS 286

QY 295 EGKI-KSGRSITIELVPXNPKIIDRIKEIQNPVFLVGFKAETSKEKLIBEGKQIERAKA 353
DB 287 ELKIKKASDNLINEFVKDN-DILKKGELIKKHQILVGFAAESN--DLIENAYGKLTKNL 343

QY 354 DLVVGNTL-----EAFGSEENQVILGRDFTK-ELPKMKKRELAERIMBIEKXLS 403
DB 344 DIVANDILSKDAGFASDENRVILGSDGSKLELDKMSKRKVAENLFDLLSKRS 398

RESULT 9

C70201
pantothenate metabolism flavoprotein dff homolog BB0812 - Lyme disease spirochete
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: C70201
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White,
son, D.; Peterson, J.; Kurlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugler,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horat, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: C70201
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-390 <LSE>
A:Cross-references: UNIPROT:O51752; GB:AE001179; GB:AE000783; NID:g2688738; PIDN:AAAC6714
A:Experimental source: strain B31
C:Superfamily: pantothenate metabolism flavoprotein dff

Query Match 27.2%; Score 543.5; DB 2; Length 390;
Best Local Similarity 36.3%; Pred. No. 2.9e-25;
Matches 144; Conservative 81; Mismatches 135; Indels 37; Gaps 14;

QY 19 KKVXVXXPGSIAAL-DVKACEGLIRHGAHVHVMSEAATKIHPYAWNLPTGNPVITEIT 77
DB 5 KHLIGICGGIASYKSVIVSLVGLGVKVKVIMTQNAKTFITPLTLETISKNIITNLW 64

QY 78 GF-----IEHVELAGEHENKADLILVCPATANTISKIACGIDDPVTVTVTTAFPHIPMI 133
DB 65 DLDHNEVEHIKIA----KWAHLILVIPATYNTISKIAGIADDAALTTIASAS--TAPTYF 118

QY 134 APAMHETMYRHPVIRENTERLKLGVFIEGRIEGR-----RAKVASIDEIVYRVIK 185
DB 119 AIAAMNIMYSPILKENIKKLKLTNYKFTIEP--DKGLACSSNALGRKNKDKIKIILN 176

QY 186 KLHKKT-LEGKRVLTAGATREYIDPIRFITNASSGKMGVALAEADFRGA-VTLI---R 240
DB 177 EFNQKDYLNKKNKILITASRTELDIDPIRVFSNTSTGKMGFCLAQEAVALGAQVTTITGPT 236

QY 241 TKGSVKAFRIKIKLVETVBEMLSAENELRSKKYDVVIMAAVSDFRPKIKAEKIKS 300
DB 237 NENDPEGVNIIKIKTAMEMYKALKIY-----NKFETIIGAAVADFPKGFHFNISKIK 290

QY 301 GR-SITIELVPXNPKIIDRIKEIQ-PNVFLVGFKAETSKEKLIBEGKQIERAKADLVV 357
DB 291 KKNINLYIKLV-KNPDIIOHIGNHKNQKQIVIGFCAENSK-NLIQKAKELKKKKNLDFFI 348

QY 358 GNTLEAFGSEENQVILGRDFTKELPKMKKRELAERI 394

Db	63	PDEKSSVVVAHIDL	---DWADLVVAPATANLKGKMGANGIADDMWVTTLAT--EAPV	116
Qy	132	MIAPAMHETMYRHP	IVRENIERLKKLGVFEFGPRIEE-----GRAKVASIDEIVRV	183
Db	117	WVAPAMNVHMIQHP	AVIRNINRLYADGVRFIEP--EEGYLACGYVGRGLEEPEKIVLRI	174
Qy	184	IK--KLHKKTLEGG	KVLVTAGATREVIDPIRITNASSGKMGVALAEADFRGA-VTLIR	240
Db	175	AEFFQEDKLLQGG	KVLTAGATREKLDPVRFTHSTGKMGFSAESAAHGANVTILIT	234
Qy	241	TGKS-----VKA	FRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAAVSDFRPKKA	294
Db	235	TSKTLVPVPGVEA	-----VIVESADEMYQAVLE--RKAEQHIFWMTAAVADYTPANVS	285
Qy	295	EKKI-KSGRSITIE	LPVXNPKIIDRIKEIQPNVFLGVFKAETSKLEBEGKQRIERAKA	353
Db	286	EQKIKKPGDFTIE	MKRTKDILLEGQNKTDQOVVIGFAAET--ENLETNALKKLTSKNA	343
Qy	354	DLVVGNTLE----	AFGSEENQVVLIGRDFTEK-LPKMKKRELAERIWDIEKKL	402
Db	344	DMIVANNISEAG	AGSGDNTNIVTYRKOGSNEALPLDKKEVAEHIEEAANFL	397
RESULT 13				
F83963				
flavoprotein dfp [imported] - Bacillus halodurans (strain C-125)				
C:Species: Bacillus halodurans				
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004				
C:Accession: F83963				
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira				
Nucleic Acids Res. 28, 4317-4331, 2000				
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and				
A:Reference number: A83650; MUID:20512582; PMID:11058132				
A:Accession: F83963				
A:Status: preliminary				
A:Molecule type: DNA				
A:Residues: 1-404 <STO>				
A:Cross-references: UNIPROT:Q9K9Y4; GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA062				
A:Experimental source: strain C-125				
C:Genetics:				
A:Gene: dfp				
C:Superfamily: pantothenate metabolism flavoprotein dfp				
Query Match 24.7%; Score 494.5; DB 2; Length 404;				
Best Local Similarity 33.5%; Pred. No. 2.5e-22;				
Matches 138; Conservative 84; Mismatches 149; Indels 41; Gaps 15;				
Qy	16	LVGKIVXXPGSIA	ALDVKA-CEGLIRHGAEVHVMSEAATKIHPYAMNLTGNPVIT	74
Db	2	LQGRVVLVSGGIA	AFKSAAFASKLVQAGAEVQVVMTEGAKKFTPLTFQALTRHPYVD	61
Qy	75	EI-----TGF	IEHVELAGEHENKADLILVCPATANTISKIACGIDDPVTVTVTTFAPPHI	129
Db	62	DFSEPDPEIAHI	QILA----DWADVITATPATANLIGLANGVADDMLSLTLLAT--KA	115
Qy	130	PMIAPAMHETMYR	HPVIRENTERLKKLGVFEFGPRIEE-----GRAKVASIDEIVRV	183
Db	116	PVLPAMNVNMYE	HPAVQRNQMLAKDGYRLLEPGAGYACGWIGKGRMPPEDLIKTI	175
Qy	184	IKKLH-----	KKTLEGRVLVTAGATREYIDPIRITNASSGKMGVALAEAE--DFRGAVTL	238
Db	176	--EVHPTPSPSL	AGKVVITAGPTQETIDPIRFTNRSRGGKMGVALTKAARDFGGNVTL	233
Qy	239	IRTKGSV-KAFRI	RKIKLVETVEEMLSAIENELRSKKYDVVIMAAAVSDFRPKKAEQK	297
Db	234	ISGPTSLERPDG	SVNVVK--SAQDMYEAVLAEFSDA--DVWIKTAAVADYPRVVHEQK	289
Qy	298	IKSGRSITIELVP	XNPKIIDRIKEIQPNVFLGVFKAETSKLEBEGKQRIERAKADLVV	357
Db	290	VKKKGDWDVIEL	ERTVDILKLGKKESQFLVGFAAES--QEVETVYQKKLKEKNADMIV	347
Qy	358	GNTL-----	EAFGSEENQV-VLIGRDFTEKLPKMKKRELAER----IWDIEIK	400

Db	348	ANNVTEGAGQTDT	NRVTYVFQGDVKKPLMTKDEVAHRLIMMISQLEK	399
RESULT 14				
A11302				
pantothenate metabolism flavoprotein homolog lmo1825 [imported] - Listeria monoc				
C:Species: Listeria monocytogenes				
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004				
C:Accession: A11302				
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker				
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.				
Science 294, 849-852, 2001				
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Me				
ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,				
A:Title: Comparative genomics of Listeria species.				
A:Reference number: AB1077; MUID:21537279; PMID:11679669				
A:Accession: A11302				
A:Status: preliminary				
A:Molecule type: DNA				
A:Residues: 1-399 <GLA>				
A:Cross-references: UNIPROT:Q8Y674; GB:NC_003210; PIDN:CAC99903.1; PID:g16411279; GSPDB:				
A:Experimental source: strain EGD-e				
C:Genetics:				
A:Gene: lmo1825				
C:Superfamily: pantothenate metabolism flavoprotein dfp				
Query Match 24.7%; Score 494; DB 2; Length 399;				
Best Local Similarity 34.8%; Pred. No. 2.6e-22;				
Matches 144; Conservative 70; Mismatches 152; Indels 48; Gaps 16;				
Qy	18	GKIVXXPGSIA	ALDVKACBG-LIRHGAEVHVMSEAATKIHPYAMNLTGNPVITEI	76
Db	3	GKNILLAVSGGIA	VYKVAVALTSKLTQAGANVKVMHTAHAEFVPLSPQVLKNDVYVDT	62
Qy	77	-----TGF	IEHVELAGEHENKADLILVCPATANTISKIACGIDDPVTVTVTTFAPPHI	131
Db	63	FDEKSSVVVAHID	LA----DWADLVVAPATANVIGKMGANGIADDMWVTTLAT--EAPV	116
Qy	132	MIAPAMHETMYR	HPVIRENTERLKKLGVFEFGPRIEE-----GRAKVASIDEIVRV	183
Db	117	WVAPAMNVHMIQ	HPAVIRNINRLYADGVRFIEP--EEGYLACGYVGRGLEEPEKIVLRI	174
Qy	184	IK--KLHKKTLEG	KVLVTAGATREVIDPIRITNASSGKMGVALAEADFRGA-VTLIR	240
Db	175	AEFFQEDKLLQGG	KVLTAGATREKLDPVRFTHSTGKMGFSAESAAHGANVTILIT	234
Qy	241	TGKS-----	VKAFRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAAVSDFRPKKA	294
Db	235	TSKALVPVPGVEA	-----IIVESAEEMHQAV-NE-RKVSQDIFWMTAAVADYTPAQVS	285
Qy	295	EKKI-KSGRSITIE	LPVXNPKIIDRIKEIQPNVFLGVFKAETSKLEBEGKQRIERAKA	353
Db	286	DQKIKKPGDFTI	AMKRTKDILLEGQHTSKTSQVVGFAAET--ENVEANARKKLTSKNA	343
Qy	354	DLVVGNTLE----	AFGSEENQVVLIGRDFTEK-LPKMKKRELAERIWDIEKKL	402
Db	344	DMIVANNISEAG	AGSGDNTNIVTYRKOGSSEALPILDKKEVAEHIEEAANFL	397
RESULT 15				
A82351				
DNA/pantothenate metabolism flavoprotein VC0215 [imported] - Vibrio cholerae (strain N1				
C:Species: Vibrio cholerae				
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004				
C:Accession: A82351				
R:Heidelberg, J.F.; Easen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.				
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, I				
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.				
Nature 406, 477-483, 2000				
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.				
A:Reference number: A82035; MUID:20406833; PMID:10952301				

Search completed: July 2, 2005, 12:25:15
Job time : 25.0698 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:17:13 ; Search time 8.93023 Seconds
(without alignments)
1680.787 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPDKIRKELIIEPFSEE.....PYRGNVQSGSTRLAFSKRKKL 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	768	95.3	156	E71216	dCTP deaminase (EC
2	703	87.2	154	G75030	probable dCTP deam
3	243	30.1	180	A70439	deoxycytidine triph
4	242.5	30.1	177	H81695	probable dCTP deam
5	242	30.0	173	S28382	probable dCTP deam
6	232.5	28.8	181	D72724	probable dCTP deam
7	216	26.8	191	T36613	deoxycytidine triph
8	207.5	25.7	206	A90160	dCTP deaminase (EC
9	199	24.7	200	A69114	dCTP deaminase (EC
10	198	24.6	190	B70526	probable deoxycyti
11	197.5	24.5	190	H87222	probable dCTP deam
12	189.5	23.5	186	A81272	deoxycytidine triph
13	184.5	22.9	195	G84184	dCTP deaminase (EC
14	181	22.5	190	D64566	dCTP deaminase (EC
15	180	22.3	188	E71860	dCTP deaminase (EC
16	177	22.0	193	A10185	dCTP deaminase (EC
17	174	21.6	206	C84942	dCTP deaminase (EC
18	172	21.3	188	A83210	probable deoxycyti
19	167.5	20.8	193	AF0770	dCTP deaminase (EC
20	165	20.5	193	H90987	2'-deoxycytidine 5
21	165	20.5	193	C85833	2'-deoxycytidine 5
22	164.5	20.4	191	G82765	deoxycytidine triph
23	163	20.2	193	A42940	dCTP deaminase (EC
24	159	19.7	163	H72759	probable dCTP deam
25	155	19.2	188	D81149	deoxycytidine triph
26	154	19.1	204	F64353	dCTP deaminase (EC
27	153	19.0	150	D69081	deoxycytidine 5-tri
28	150.5	18.7	190	D81717	deoxycytidine triph
29	149.5	18.5	190	B71565	probable dCTP deam

30	149.5	18.5	195	1	A64050	dCTP deaminase (EC
31	148	18.4	190	2	C97712	hypothetical prote
32	145	18.0	188	2	E71715	probable dCTP deam
33	138	17.1	190	2	F72084	dCTP deaminase (EC
34	138	17.1	190	2	F86539	dCTP deaminase (im
35	137.5	17.1	199	2	AB1815	dCTP deaminase (im
36	137	17.0	172	2	T44356	probable dCTP deam
37	135.5	16.8	193	1	S75588	dCTP deaminase (EC
38	134	16.6	173	2	A96903	deoxycytidine triph
39	128.5	15.9	157	2	C97049	deoxycytidine 5-tri
40	127.5	15.8	166	2	A99475	deoxycytidine triph
41	126	15.6	168	2	C69388	probable dCTP deam
42	125.5	15.6	1145	1	GNLJEV	HIV-1 retropepsin
43	125.5	15.6	1146	1	GNLJ22	HIV-1 retropepsin
44	125.5	15.6	1146	1	GNLJEV	HIV-1 retropepsin
45	121	15.0	147	2	D95002	hypothetical prote

ALIGNMENTS

RESULT 1

E71216
dCTP deaminase (EC 3.5.4.13) PH1997 [similarity] - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C:Accession: E71216
R:Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki,
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguch
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: AV1000; MUID:98344137; PMID:9679194
A:Accession: E71216
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-156 <KAW>
A:Cross-references: UNIPROT:O57706; GB:AP0000007; NID:g3236134; PIDN:BAA31124.1; PID:g32
A:Experimental source: strain OF3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1997
C:Keywords: hydrolase

Query Match	95.3%	Score	768	DB	2	Length	156
Best Local Similarity	93.6%	Pred. No.	3.6e-65				
Matches	146	Conservative	4	Mismatches	6	Indels	0
Gaps	0						
Qy	1	MLLPDKIRKELIIEPFSEESLQPAGYDLRVGREAFVKGKLDVEKEGKVPPREYALI	60				
Db	1	MLLPDKIRKELIIEPFSEESLQPAGYDLRVGREAFVKGKLDVEKEGKVPPREYALI	60				
Qy	61	LTLEIRKLPDDVGMKIRSSLAEGVIGSFAWVDPGWDGNLTMLYNASNEPVELRYGE	120				
Db	61	LTLEIRKLPDDVGMKIRSSLAEGVIGSFAWVDPGWDGNLTMLYNASNEPVELRYGE	120				
Qy	121	RFVQIAFIRLEGPARNPYRGNYQSGSTRLAFSKRKKL	156				
Db	121	RFVQIVFIRLEPPRPNYSGNYQSGSTRLAFSKRKKL	156				

RESULT 2

G75030
dCTP deaminase (EC 3.5.4.13) PAB1164 [similarity] - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: G75030
R:Anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A:Reference number: A75001
A:Accession: G75030
A:Status: preliminary
A:Molecule type: DNA

A;Residues: 1-154 <KAW>
A;Cross-references: UNIPROT:Q9UXS8; GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB5068
A;Status: Preliminary
A;Molecule type: DNA
C;Genetics:
A;Gene: dcd; PAB1164
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

Query Match 87.2%; Score 703; DB 2; Length 154;
Best Local Similarity 85.6%; Pred. No. 4.7e-59;
Matches 131; Conservative 13; Mismatches 59; Indels 0; Gaps 0;
Qy 1 MLLPDKIRKEILIPFSEESLOPAGYDLRVGREAFVKGKIDIVEKEGKVIPPPEYALI 60
Db 1 MLLPDKIRKEILIPFSEESLOPAGYDLRVGREAYIOGKIDIVEKEGKVIPPPEYALI 60
Qy 61 LTLERIKLPDDVMGDMKIRSSLRARSGVIGSFAWVDPGWDGNLTLMLYNASNEPVELRYGE 120
Db 61 LTLERIKLPDDVMGDMKIRSSLRARSGVLGSAWVDPGWDGNLTLMLYNASEKEVILRYKE 120
Qy 121 RFOVQAFIRLEGPARNPYRGNYQGSSTRLAFSKR 153
Db 121 RFOVQAFIRLEGAPAKNPYRGNYQGSRRIVLSKR 153

RESULT 3
A70439
probable dCTP deaminase (EC 3.5.4.13) dcd [similarity] - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C;Accession: A70439
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ovi
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: A70439
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-180 <AQF>
A;Cross-references: UNIPROT:O67539; GB:AE000747; NID:g2983944; PIDN:AAC07499.1; PID:g298
A;Experimental source: strain VF5
C;Genetics:
A;Gene: dcd
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

Query Match 30.1%; Score 243; DB 2; Length 180;
Best Local Similarity 37.7%; Pred. No. 1.2e-15;
Matches 61; Conservative 30; Mismatches 53; Indels 18; Gaps 5;
Qy 1 MLLPDKIRK-----EILIPFSEESLOPAGYDLRVGRE-APVKGK-LIDVEKEGKV-- 51
Db 1 MILSDRSIRELIEKELKVEPYPESHVQCSSLDLRLGNQIALYEGEGVIDVKKGTKGVRI 60
Qy 52 -----IPREYALILTLERIKLPDDVMGDMKIRSSLRARSGV-IGSFAWVDPGWDGNL 102
Db 61 LEFEYFDMPKQFLATLLEYISLPPYVATFVGRSSLRGLGFENAGWVDAGFEGQI 120
Qy 103 TMLYNASNEPVELRYGSRFVQAFIRLEGPARNPYRGNYQG 144
Db 121 TLELFNANDRPIRLYRGHRIQCLVFARLDRPERYYSCKYKG 162

RESULT 4
H83695
deoxycytidine triphosphate deaminase BH0368 [imported] - Bacillus halodurans (strain C-1
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C;Accession: H83695
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: H83695
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-177 <STO>
A;Cross-references: UNIPROT:Q9KFV3; GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BAB040
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0368
C;Superfamily: dCTP deaminase

Query Match 30.1%; Score 242.5; DB 2; Length 177;
Best Local Similarity 37.2%; Pred. No. 1.4e-15;
Matches 58; Conservative 30; Mismatches 47; Indels 21; Gaps 6;
Qy 10 KEILIPFSEESLOPAGYDLRVGREAFV-----KGLIDIVEK-----EGKVVIPP 54
Db 15 KELBITLTEREQIQPASVDLRLGPH-FVTIDDSKEAVISFERPIRYREWTTSDETIVLPP 73
Qy 55 REYALILTLERIKLPDDVMGDMKIRSSLRARSGV-IGSFAWVDPGWDGNLTLMLYNASNEP 113
Db 74 HTFLATTMETVKLPNHLTAFVEGRSSVGRGLGFIONAGWVDPGFNGQITLLELFNANRLP 133
Qy 114 VELRYGSRFVQAFIRLEGPARNPYRGNY---QGST 146
Db 134 IELPIGRRICQLVFAEVTGEVA-PYQGYLPQKGAT 168

RESULT 5
S26382
probable dCTP deaminase (EC 3.5.4.13) [similarity] - Desulfurolobus ambivalens
N;Alternate names: hypothetical protein 3 lig-region
C;Species: Desulfurolobus ambivalens
C;Date: 22-Nov-1993 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C;Accession: S26382
R;Kletzin, A.
Nucleic Acids Res. 20, 5389-5396, 1992
A;Title: Molecular characterisation of a DNA ligase gene of the extremely thermophilic a
A;Reference number: S26382; MUID:93065206; PMID:1437556
A;Accession: S26382
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-173 <KLE>
A;Cross-references: UNIPROT:Q02103; EMBL:X63438; NID:g40784; PIDN:CAA45033.1; PID:g40785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1991
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

Query Match 30.0%; Score 242; DB 2; Length 173;
Best Local Similarity 35.9%; Pred. No. 1.5e-15;
Matches 56; Conservative 30; Mismatches 56; Indels 14; Gaps 3;
Qy 12 ILIPFSEESLOPAGYDLRVG-----REAFVKGK-----LIDVEKEGKVIPPPEYA 58
Db 17 IVISPLTQDTIRENGVDLRVGGELARFKKTDIYEDGKDRSFYEIEKGDFIIYPNHSV 76
Qy 59 LILTLERIKLPDDVMGDMKIRSSLRARSGVIGSFAWVDPGWDGNLTLMLYNASNEPVELRY 118
Db 77 LLVTEEVYKLPNDVMAFVNLKRSFARLGLFVPPTVDAGFEGQITIEVL-GSAFPVKIR 135
Qy 119 GERFVQAFIRLEGPARNPYRGNYQGSSTRLAFSKR 154
Db 136 GTRFLHLIFARTLTPVENPYHGKYQGQGVTLPRK 171

RESULT 6
D72724
probable dCTP deaminase (EC 3.5.4.13) APE0333 [similarity] - Aeropyrum pernix (strain K1
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: D72724
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: D72724
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-181 <RAW>
A;Cross-references: UNIPROT:Q9YFA8; DDBJ:AP000059; NID:95103911; PIDN:BAAY9288.1; PID:95103911
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0333
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

Query Match 28.8%; Score 232.5; DB 2; Length 181;
Best Local Similarity 31.8%; Pred. No. 1.2e-14;
Matches 55; Conservative 38; Mismatches 57; Indels 23; Gaps 4;
QY 1 MLLPDWKIRK-----BILIEPFSEESLQAGYDLRVGREAFVK-----G 39
Db 4 LILSDRDLKYLKESWIKIQLREDTIRENGVDLRVGNBIARPKTKDKIPDPNPPSPF 62
QY 40 KLIDVEKGVIPPREVALIILTLERIKLPDDVMGDMKIRSSLAAREGVIGSFAWVDPGWD 99
Db 63 EFVECGEGDEIIVGPEHMLLHTQEVIRLPYVAGLVNLRSTWARTGIIPATVVDAGFE 122
QY 100 GNLTLMLYNASNEPVELRYGERFVOIAFIRLEGPARNPYRGVNYQSTRLAFSK 152
Db 123 GQLTIWV-GSGFPVKLYPGDRFLHLVLVKLQSPANNPYRGVYQGGVRLPK 174

RESULT 7
T36613
probable dCTP deaminase (EC 3.5.4.13) SCH35.46 [similarity] - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T36613
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1999
A;Reference number: Z21610
A;Accession: T36613
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-191 <OLI>
A;Cross-references: UNIPROT:Q9X8W0; EMBL:AL078610; PIDN:CAB44381.1; GSPDB:GN000070; SCOE
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOE03:SCH35.46
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

Query Match 26.8%; Score 216; DB 2; Length 191;
Best Local Similarity 31.4%; Pred. No. 4.7e-13;
Matches 53; Conservative 32; Mismatches 60; Indels 24; Gaps 3;
QY 1 MLLPDWKIRKEI-----LIEPFSEESLQAGYDLRVGR-----EAFVK 38
Db 1 MLLSDKIDRAEIDNGRVIDPDDSMVQPSIDVRLDRYFRVFNHRYPHIDPSVEQVDL 60
QY 39 GKLIDVEKGVIPPREVALIILTLERIKLPDDVMGDMKIRSLAREGVI--GSFAWVDP 96
Db 61 TRLVEPEGDEPFIHGFGLVASTYEVWSPDDLASRLGSKSLGLGLVTHSTAGFIDP 120
QY 97 GWDGNLTLMLYNASNEPVELRYGERFVOIAFIRLEGPARNPYRGVNYQGS 145
Db 121 GFGSHVTLSNLATLPKLVQPMQKIGQLCLFRLTSPAEPHYGSEYRYS 169

RESULT 8
A90160
deoxycytidine triphosphate deaminase (dcD-1) [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 17-May-2002

C;Accession: A90160
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A90139
A;Accession: A90160
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-206 <KUR>
A;Cross-references: GB:AE006641; NID:gl3813326; PIDN:AAK40536.1; GSPDB:GN00155
C;Genetics:
A;Gene: dcd-1
C;Superfamily: dCTP deaminase

Query Match 25.7%; Score 207.5; DB 2; Length 206;
Best Local Similarity 33.5%; Pred. No. 3.2e-12;
Matches 57; Conservative 30; Mismatches 64; Indels 19; Gaps 5;
QY 1 MLLPDWKI-----RKEILIEPFSEESLQAGYDLRVGRE--AFVK-GKLID----- 43
Db 23 MILSDRDLKYLKESWIKIQLREDTIRENGVDLRVGNBIARPKTKDKIPDPNPPSPF 82
QY 44 -VEKEGVIPPREVALIILTLERIKLPDDVMGDMKIRSLAREGVIGSFAWVDPGWDNL 102
Db 83 QTEKGEEFTIQPYEHVLLTTEYIELNDVMAFVNLRTFARLGLFIPPTIVDAGFGQV 142
QY 103 TLMLYNASNEPVELRYGERFVOIAFIRLEGPARNPYRGVNYQSTRLAFSK 152
Db 143 TIEVW-GSFPVKLRKSTRFIHLIFARTLTPVEYQGGYQGGVTLPK 191

RESULT 9
A69114
dCTP deaminase (EC 3.5.4.13) MTH1847 [similarity] - Methanobacterium thermoautotrophicum
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 31-Mar-2000
C;Accession: A69114
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: A69114
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-200 <MTH>
A;Cross-references: GB:AE000937; GB:AE000666; NID:g2622974; PIDN:AAB86313.1; PID:g26229
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1847
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

Query Match 24.7%; Score 199; DB 2; Length 200;
Best Local Similarity 31.6%; Pred. No. 2e-11;
Matches 59; Conservative 29; Mismatches 63; Indels 36; Gaps 7;
QY 2 LLLPDWKIRKE-----LIEPFSEESLQAGYDLRVGREAFVKGLI----- 42
Db 6 ILSDRDLKRYIEEGLTIDPLDPERIQIPSSVDLRIGNE--FKGPRVIRKPCIDPKDPS 63
QY 43 DVE-----KEGKVIPPREVALIILTLERIKLPDDVMGDMKIRSLAREGVI--GSFA 92
Db 64 DIESYMETPHVEDGPFIIHPGFALATTHEYIALPEDLVARVEGRSSIGRLGTHMHTAG 123
QY 93 WYDPGWDGNLTLMLYNASNEPVELRYGERFVOIAFIRLEGPARNPYRGVNYQGS 147
Db 124 YIDPFGHRTITLEISNIGKMPVALYPRQVVCQIVFETMTSPAERFPGHSPRDSKYIGQTR 183
QY 148 LAFSKRK 154

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 2, 2005, 08:22:22 ; Search time 3629 Seconds
(without alignments)
2082.947 Million cell updates/sec

Title: US-08-957-709a-71

Perfect score: 806

Sequence: 1 MLPLPWKIRKEILIEPFSEE.....PYRGNVQGSTRLAFKRKKL 156

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2.1/USPTO.spool/US08957709/runat_01072005.154631.26005/app.query.fasta_1.327
-DB=GenEmbl -QFMT=fastap -SUFFIX=rgc -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US08957709 @CNG 1 1 4200 @runat 01072005.154631.26005 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	806	100.0	471	1	AY066005 Pyrococcus
2	806	100.0	471	1	AY443493 Pyrococcus
3	806	100.0	471	6	BD133238 Pyrococcus
4	806	100.0	471	6	AR535070 Sequence

C	5	806	100.0	13064	1	AE010292	AE010292 Pyrococcus
	6	788	95.3	253505	1	AP000007	AP000007 Pyrococcus
	7	703	87.2	65118	6	BD430793_17	Continuation (18 o
	8	703	87.2	265118	1	CNSPAX06	AJ248288 Pyrococcus
	9	703	87.2	265118	6	AX041922	AX041922 Pyrococcus
	10	543	67.4	740	6	BD133228	BD133228 Pyrococcus
	11	543	67.4	740	6	AR180965	AR180965 Pyrococcus
	12	543	67.4	740	6	AR207520	AR207520 Pyrococcus
	13	543	67.4	740	6	AR225643	AR225643 Pyrococcus
	14	543	67.4	740	6	AR535060	AR535060 Pyrococcus
	15	257.5	31.9	338100	7	TAGID2	AL445064 Thermopla
	16	245.5	30.5	75294	7	AJ783769	AJ783769 Sulfolobu
	17	244	30.3	35450	14	SVI344259	AJ344259 Sulfolobu
	18	243	30.1	14286	1	AE000747	AE000747 Aquifex a
	19	242.5	30.1	296950	1	AP001508	AP001508 Bacillus
	20	242	30.0	3382	1	DALIG	X63438 D.ambivalen
	21	242	30.0	3382	6	A78674	A78674 Sequence 5
	22	237	29.4	32308	14	AF022221	AF022221 Archaeal
	23	237	29.4	32308	14	SVI414696	AJ414696 Sulfolobu
	24	236	29.3	10898	1	AE009852	AE009852 Pyrococcus
	25	233	28.9	543	6	BD180240	BD180240 Highly th
	26	233	28.9	247910	1	AE017307	AE017307 Thermus t
	27	233	28.9	263050	1	AP000981	AP000981 Sulfolobu
	28	232.5	28.8	167000	1	AP000059	AP000059 Aeropyrum
	29	232	28.8	110000	1	AE017261_01	Continuation (2 of
	30	227.5	28.2	110000	1	AE017199_2	Continuation (3 of
	31	220.5	27.4	239340	1	AP000993	AP000993 Thermopla
	32	220	27.3	10375	1	AE010327	AE010327 Methanopy
	33	218.5	27.1	302174	1	AE017241	AE017241 Mycobacte
	34	214	26.6	567	6	BD165341	BD165341 Novel pol
	35	214	26.6	567	6	AX123224	AX123224 Sequence
	36	214	26.6	309400	6	AX127153	AX127153 Sequence
	37	214	26.6	325651	1	AP005283	AP005283 Corynebac
	38	214	26.6	349115	1	BX927156	BX927156 Corynebac
	39	214	26.6	349980	6	AX127152	AX127152 Sequence
	40	212	26.3	309050	1	SC0939117	AL939117 Streptomy
	41	210	26.1	299925	1	AP005039	AP005039 Streptomy
	42	209.5	26.0	110000	1	AP006618_57	Continuation (58 o
	43	207.5	25.7	11922	1	AE006657	AE006657 Sulfolobu
	44	205.5	25.5	302938	1	BX251412	BX251412 Tropherym
	45	205.5	25.5	324227	1	AE016852	AE016852 Tropherym

ALIGNMENTS

RESULT 1	AY066005	Pyrococcus furiosus dutPase gene, complete cds.	471 bp	DNA	linear	BCT 28-JAN-2002
LOCUS	AY066005	Pyrococcus furiosus dutPase gene, complete cds.				
DEFINITION	AY066005	Pyrococcus furiosus dutPase gene, complete cds.				
ACCESSION	AY066005.1	GI:18389121				
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
MEDLINE						
PUBMED						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
Source						

CDS	1..471 /codon_start=1 /transl_table=11 /product="dUTPase" /protein_id="AAL47572.1" /db_xref="GI:18389122" /translation="MLLPDWKIRKEILPEPSEESLOPAGYDLRVGREAFVKGKLLDV EKEGKVIPPPEYALILTLERIKLPDDVMGDMKIRSSLAESGVIGSFAMVDPGWGDNL TMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL"	
ORIGIN		
Alignment Scores:	1.04e-80	Length: 471
Pred. No.:	806.00	Matches: 156
Score:	100.00%	Conservative: 0
Percent Similarity:	100.00%	Mismatches: 0
Best Local Similarity:	100.00%	Indels: 0
Query Match:	100.00%	Gaps: 0
DB:	1	
US-08-957-709A-71 (1-156) x AY066005 (1-471)		
QY	1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu 20	
Db	1 ATGCTACTTCCAGACTGGAAATCAGAAAAGAAATACTATAGAGCCATTTCTGAAGAA 60	
QY	21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40	
Db	61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGCAGAGAGGCTTTTGTAAAGGGGAAA 120	
QY	41 LeuIleAspValGluLysGluGlyValValIleProProArgGluTyrAlaLeuIle 60	
Db	121 TTAATCGACGTGGAAAGAGAAAGAAAGTCGTTATTCCTCCAAAGGGAATACGCCCTTAATC 180	
QY	61 LeuThrLeuGluArgLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80	
Db	181 CTAACCTCTCAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240	
QY	81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly 100	
Db	241 AGTTTAGCAAGAGAGAGGGGTTATTGGTTCTTTGGTTGGTCCAGCCAGGATGGATGGA 300	
QY	101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120	
Db	301 AACTTAACTAATGCTCTACATGCTCAATGAAACCTGTGCAATTAAGATATGAGAG 360	
QY	121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140	
Db	361 AGATTGTGCAGATCGATTTATAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAGGA 420	
QY	141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLysLeu 156	
Db	421 AACTATCAGGGGAGCAAGGTTAGCGTTTTCANAGAGAGAAAGAACTC 468	
RESULT 2		
AY443493	471 bp	DNA linear BCT 17-NOV-2003
LOCUS	Pyrococcus woesei dUTPase gene, complete cds.	
DEFINITION		
ACCESSION	AY443493	
VERSION	AY443493.1	GI:38261503
KEYWORDS		
SOURCE	Pyrococcus woesei	
ORGANISM	Pyrococcus woesei	
REFERENCE	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus.	
AUTHORS	1 (bases 1 to 471)	
TITLE	Dabrowski, S. and Kjaer Ahring, B. Cloning, expression, and purification of the His6-tagged hyper-thermostable dUTPase from Pyrococcus woesei in Escherichia coli: application in PCR	
JOURNAL	Protein Expr. Purif. 31 (1), 72-78 (2003)	
MEDLINE	22844507	
PUBMED	12963343	
REFERENCE	2 (bases 1 to 471)	
AUTHORS	Dabrowski, S. and Ahring, B. K.	
TITLE	Direct Submission	
JOURNAL	Submitted (20-OCT-2003) Biocentrum-DTU, Technical University of Denmark, Soltofts Plads 227, Kgs. Lyngby 2800, Denmark	
FEATURES	Location/Qualifiers	
source	1..471	
CDS	/organism="Pyrococcus woesei" /mol_type="genomic DNA" /db_xref="taxon:2262" 1..471 /codon_start=1 /transl_table=11 /product="dUTPase" /protein_id="AAR15897.1" /db_xref="GI:38261504" /translation="MLLPDWKIRKEILPEPSEESLOPAGYDLRVGREAFVKGKLLDV EKEGKVIPPPEYALILTLERIKLPDDVMGDMKIRSSLAESGVIGSFAMVDPGWGDNL TMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL"	
ORIGIN		
Alignment Scores:	1.04e-80	Length: 471
Pred. No.:	806.00	Matches: 156
Score:	100.00%	Conservative: 0
Percent Similarity:	100.00%	Mismatches: 0
Best Local Similarity:	100.00%	Indels: 0
Query Match:	100.00%	Gaps: 0
DB:	1	
US-08-957-709A-71 (1-156) x AY443493 (1-471)		
QY	1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu 20	
Db	1 ATGCTACTTCCAGACTGGAAATCAGAAAAGAAATACTATAGAGCCATTTCTGAAGAA 60	
QY	21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40	
Db	61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGCAGAGAGGCTTTTGTAAAGGGGAAA 120	
QY	41 LeuIleAspValGluLysGluGlyValValIleProProArgGluTyrAlaLeuIle 60	
Db	121 TTAATCGACGTGGAAAGAGAAAGAAAGTCGTTATTCCTCCAAAGGGAATACGCCCTTAATC 180	
QY	61 LeuThrLeuGluArgLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80	
Db	181 CTAACCTCTCAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240	
QY	81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly 100	
Db	241 AGTTTAGCAAGAGAGAGGGGTTATTGGTTCTTTGGTTGGTCCAGCCAGGATGGATGGA 300	
QY	101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120	
Db	301 AACTTAACTAATGCTCTACATGCTCAATGAAACCTGTGCAATTAAGATATGAGAG 360	
QY	121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140	
Db	361 AGATTGTGCAGATCGATTTATAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAGGA 420	
QY	141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLysLeu 156	
Db	421 AACTATCAGGGGAGCAAGGTTAGCGTTTTCANAGAGAGAAAGAACTC 468	
RESULT 3		
BD133238	471 bp	DNA linear PAT 18-SEP-2002
LOCUS	Pyromeraze enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF proteins, and methods for purifying and identifying them.	
DEFINITION		
ACCESSION	BD133238	
VERSION	BD133238.1	GI:23228183
KEYWORDS	JP 2002505572-A/34.	
SOURCE	unidentified	
ORGANISM	unidentified	

unclassified.
1 (bases 1 to 471)
REFERENCE Hogrefe,H. and Hansen,C.J.
AUTHORS Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
TITLE isolated PEF proteins, and methods for purifying and identifying
JOURNAL Patent: JP 2002505572-A 34 19-FEB-2002;
STRATAGENE

COMMENT
OS Unidentified
PN JP 2002505572-A/34
PD 19-FEB-2002
PF 20-MAR-1998 JP 1998545828
PR 21-MAR-1997 US 08/822774,24-OCT-1997 US 08/957709 PI
PC HOLLY HOGREFE, CONNIE J HANSEN
PC C12P19/34, C12Q1/68, C12N9/12
CC Strandedness: Single;
CC Topology: Linear;
CC Polymerase enhancing factor (PEF) extracts, PEF protein CC
complexes,
CC isolated PEF proteins, and methods for purifying and CC
identifying them

PH Key Location/Qualifiers
FT source 1..471
FT /organism='Unidentified'.

FEATURES
source
1..471
Location/Qualifiers
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'

ORIGIN

Alignment Scores:
Pred. No.: 1.04e-80 Length: 471
Score: 806.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-08-957-709A-71 (1-156) x BD133238 (1-471)

QY 1 MetLeuLeuProAspTrrpLysileArgLysGluLeuLeuGluProPheSerGluGlu 20
DB 1 ATGCTACTTCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAGAA 60
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 61 TCGCTCCAAACCAAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTCTTAAGGGGAAA 120
QY 41 LeuileAspValGluLysGluGlyLysValValleProProArgGluTyrAlaLeulle 60
DB 121 TTAATCGACGTGAAAGGAAAGGAAAGTCTGTTATTCCTCCAAAGGGAATACGCTTAATC 180
QY 61 LeuThrLeuGluArgLileLysLeuProAspAspValMetGlyAspMetLysileArgSer 80
DB 181 CTAACCTCGAGAGGATAAGTTGCCCGACGATGTTATGGGGATATGAAGATAGAGGC 240
QY 81 SerLeuAlaArgGluGlyValleGlySerPheAlaTrrpValAspProGlyTrrpAspGly 100
DB 241 AGTTTAGCAAGAGAGGGGTTATGCTTTCTTTTGGTTGACCCAGGATGGATGA 300
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 301 AACTTAACACTAATGCTCTACAAATGCAATGCAATGCAATTAAGATATGAGAG 360
QY 121 ArgPheValGlnleAlaPheleArgLeuGluGlyProAlaArgAsnProTrrpArgGly 140
DB 361 AGATTGTGCAGATCGCATTTATAAGGCTAGAGGGTCCGGCAAGAAACCTTTACAGAGGA 420
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLysLeu 156
DB 421 AACTATCAGGGGAGCAAGGTTAGCGTTTTCAGAGAGAAAGAACTC 468

RESULT 4

AR535070

LOCUS AR535070 471 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 70 from patent US 6734293.
ACCESSION AR535070
VERSION AR535070.1 GI:53925890
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 471)
AUTHORS Hogrefe,H. and Hansen,C.J.
TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
JOURNAL isolated PEF protein, and methods for purifying and identifying
FEATURES Patent: US 6734293-A 70 11-MAY-2004;
source Location/Qualifiers
1..471
/organism='unknown'
/mol_type='genomic DNA'

ORIGIN

Alignment Scores:
Pred. No.: 1.04e-80 Length: 471
Score: 806.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-08-957-709A-71 (1-156) x AR535070 (1-471)

QY 1 MetLeuLeuProAspTrrpLysileArgLysGluLeuLeuGluProPheSerGluGlu 20
DB 1 ATGCTACTTCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAGAA 60
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 61 TCGCTCCAAACCAAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTCTTAAGGGGAAA 120
QY 41 LeuileAspValGluLysGluGlyLysValValleProProArgGluTyrAlaLeulle 60
DB 121 TTAATCGACGTGAAAGGAAAGGAAAGTCTGTTATTCCTCCAAAGGGAATACGCTTAATC 180
QY 61 LeuThrLeuGluArgLileLysLeuProAspAspValMetGlyAspMetLysileArgSer 80
DB 181 CTAACCTCGAGAGGATAAGTTGCCCGACGATGTTATGGGGATATGAAGATAGAGGC 240
QY 81 SerLeuAlaArgGluGlyValleGlySerPheAlaTrrpValAspProGlyTrrpAspGly 100
DB 241 AGTTTAGCAAGAGAGGGGTTATGCTTTCTTTTGGTTGACCCAGGATGGATGA 300
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 301 AACTTAACACTAATGCTCTACAAATGCAATGCAATGCAATTAAGATATGAGAG 360
QY 121 ArgPheValGlnleAlaPheleArgLeuGluGlyProAlaArgAsnProTrrpArgGly 140
DB 361 AGATTGTGCAGATCGCATTTATAAGGCTAGAGGGTCCGGCAAGAAACCTTTACAGAGGA 420
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLysLeu 156
DB 421 AACTATCAGGGGAGCAAGGTTAGCGTTTTCAGAGAGAAAGAACTC 468

RESULT 5

AE010292/c
LOCUS AE010292 13064 bp DNA linear BCT 25-FEB-2002
DEFINITION Pyrococcus furiosus DSM 3638, section 167 of the complete
genome.
ACCESSION AE010292 AE009950
VERSION AE010292.1 GI:18894190
KEYWORDS
SOURCE Pyrococcus furiosus DSM 3638
ORGANISM Pyrococcus furiosus DSM 3638
Archaea; Euryarchaeota; Thermococci; Thermococcales;


```

RKLEVLVPHGRKGVKVIADGVAEAAARKLGLDVISSAELEETASSPROARKLAK
KYDFFTAEPAPKPKIGRYLGRVLPKRNKMPVVPPRLDTLPIVEKLLKTKVRIQLKN
PVVHAPVGTGKMSDEEIAENIEAVLNAILGKERSSQVKSIVTKTGPVAKIEG"
/genes="PF1993"
5520. .6539
/genes="PF1993"
5520. .6539
/note="Function Code: 13.4 Translation: Ribosomal
proteins: synthesis and modification; (rpl108)"
/codon start=1
/transl table=11
/product="LSU ribosomal protein L10E"
/protein_id="AAL82117.1"
/db_xref="GI:18894199"
/transl table=11
/product="MAHVAEWKKKEVEELANLIKSPVVALVDVSMMPAYPLSQMRRL
IRENGLLSVNTLIELAKVQELGRPELEKLYINEGGAGILVITMNPFLKYKF
LQONQPAKPAKCAKVPKDVIPAGPISLAPGPIVGQMQAMGIPARIERKGVTKIDT
VYLKAGEITPELANLALQIPELVGLDLAVVEDGIITPDLVAIDSEVINMLQ
KAYMAFLNVAIVPTPOTIEALIQKAFNAKAVAEAGYITKETISDILIGRAIPAM
LILLAQLPREDVDEKTKELLSAQVSAQVVEEKKKEEKEEKEEVEEVEEALAGL
SALFG"
6586. .6909
/genes="PF1994"
6586. .6909
/genes="PF1994"
/note="Function Code: 13.4 Translation: Ribosomal
proteins: synthesis and modification; (rpl12A)"
/codon start=1
/transl table=11
/product="LSU ribosomal protein L12A"
/protein_id="AAL82118.1"
/db_xref="GI:18894200"
/transl table=11
/product="MEYVYAALLHSVKGKEEINENLKVLAQAGVEPDEARIKALVAA
LEGVINDEIVIEKAAMPVAVAAAPAGGGBEKEEKEEKEEVEEVEEVEEALAGL
SALFG"
7014. .7556
/genes="PF1995"
7014. .7556
/genes="PF1995"
/note="Function Code: 16.1 Conserved Hypothetical"
/codon start=1
/transl table=11
/product="hypothetical protein"
/protein_id="AAL82119.1"
/db_xref="GI:18894201"

Alignment Scores:
Pred. No.: 4,86e-79 Length: 13064
Score: 806.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-08-957-709A-71 (1-156) x AE010292 (1-13064)

QY 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu 20
Db 8006 ATGCTACTTCAGACTGGAATATCAGAAAGAAATCTATATAGAGCCATTTCTGAAGAA 7947
QY 21 SerLeuGluProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
Db 7946 TCGCTCAACACGACGCTTATGACCTCAGAGTGGGCAGAGAGGCTTTTGTAAAGGGGAAA 7887
QY 41 LeuIleAspValGluLysGluGlyLysValValIleProArgGluTyrAlaLeuIle 60
Db 7886 TTAATCGAGCTGGAAGGAGGAAAGTCGTTATTCCTCCAGGGAATACGCTTAATC 7827
QY 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80
Db 7826 CTAACCTCGAGGAGTAAGTTGCCCGACGATGTTATCGGGGATATGAGATAGGAGC 7767
QY 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly 100
```

```

|||||
7766 AGTTTAGCAAGAGAGGGGTTATTGTTCTTTTGGTTGGTACCAGGATGGATGGA 7707
|||||
101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
|||||
7706 AACTTAACACTTAATGCTCTACAATGCCTCAATGAACCTGTGCGAATTAAGATATGAGAG 7647
|||||
121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
|||||
7646 AGATTTTGTGCAGATCGCATTTATTAAGGCTAGAGGGTCCGCAAGAAACCTTACAGAGGA 7587
|||||
141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLysLeu 156
|||||
7586 AACTATCAGGGAGCAAGGTTAGCGTTTTCANAGAGAAAGAACTC 7539
|||||

AP000007 253505 bp DNA linear BCT 27-MAY-2004
Pyrococcus horikoshii OT3 DNA, complete genome, 1485001-1738505 nt.
position, section 7/7.
AP000007 AB009464 AB009465 AB009521 AB009522 AB009523 AB009524
AB009525 AB009526 AB009527 AB009528 AB009529 AB009530 AB009531
BA000001
AP000007.1 GI:3236134
Pyrococcus horikoshii OT3
Pyrococcus horikoshii OT3
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
REFERENCE
1
AUTHORS
Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y.,
Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y.,
Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Onfuku, Y.,
Funahashi, N., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushiida, N.,
Oguchi, A., Aoki, K., Nakamura, Y., Robb, T.F., Horikoshi, K.,
Masuchi, Y., Shizuya, H. and Kikuchi, H.
Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3
DNA Res. 5 (2), 55-76 (1998)
98344137
9679194
2 (bases 1 to 253505)
Director-General of Biotechnology Center, Tanaka, T.,
Kawarabayasi, Y. and Kikuchi, H.
Direct Submission
Submitted (11-JUN-1998) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center; 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan
(E-mail: bio@nitech.go.jp, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)
On or before Mar 17, 1999 this sequence version replaced
GI:3131896, GI:3131920, GI:3131974, GI:3131982, GI:3132032,
GI:3132063, GI:3132118, GI:3132139, GI:3132184, GI:3132233,
GI:3133142.
Kawarabayasi, Y. is officially affiliated with the National
Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
305-0046, Japan.
Robb, T. F. is at the Center of Marine Biotechnology, University of
Maryland, Baltimore, MD, USA.
Horikoshi, K. is at the Japan Marine Science and Technology Center,
Yokosuka, Kanagawa 237-0061, Japan.
Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041,
Japan.
Shizuya, H. is at the California Institute of Technology, Pasadena,
CA, USA.
The other authors are at the National Institute of Technology and
Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
All the sequence with length 100 codons or more between ATG or GTG
and stop codon are defined as CDS
Homology analysis is performed by Smith-Waterman algorithm against
GenBank and GenPept release 103; EMBL release 52.0; SwissProt
release 34.0; PIR-Protein release 54.0; and OWL release 29.5.
E-mail address for comments and questions: genomeOT3@nitech.go.jp
Restriction map, ORF organization, sequence alignment and more
```

information are available at W.W.W. site of Biotechnology Center,
URL: <http://www.bio.nits.go.jp/>.

FEATURES

source
1..253505
/organism="Pyrococcus horikoshii OT3"
/mol_type="genomic DNA"
/strain="OT3"
/db_xref="taxon:70601"
/map="1485001-1738505 nt. position"

ORIGIN

Alignment Scores:
Pred. No.: 2 648-73 Length: 253505
Score: 768.00 Matches: 146
Percent Similarity: 96.15% Conservatives: 4
Best Local Similarity: 93.59% Mismatches: 6
Query Match: 95.29% Indels: 0
DB: 1 Gaps: 0

US-08-957-709A-71 (1-156) x AP000007 (1-253505)

QY 1 MetLeuLeuProAspTrpLysLeuArgLysGluLeuLeuGluProPheSerGluGlu 20
DB 250855 ATGCTACTTCGGAGCTGGAAATAAGAAAGAAATACCTATAGAACCACTTTCTCGAAGAA 250914
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 250915 TCACCTCAACCCAGCAGGTATGATCTTAGAGTGGGTAGAGAGCTTTTGTAGTCGAAAG 250974
QY 41 LeuLeuAspValGluLysGluGlyValValLeuProArgGluTyrAlaLeuLeu 60
DB 250975 TTAATGATGTGGAAGAGGGAAGTAGTATTATCTCCAGGGAATACGCTCTAATT 251034
QY 61 LeuThrLeuGluArgLysLeuProAspValMetGlyAspMetLysLeuArgSer 80
DB 251035 CTAACCCTCGAGAGATTAAGTTACTGTATGATGTGATGGGAGATATGAAGATAAGGAGC 251094
QY 81 SerLeuAlaArgGluGlyValLeuGlySerPheAlaTrpValAspProGlyTrpAspGly 100
DB 251095 AGCTTAGCTAGAGAGGGATTTAGGTCTTTTTCCTGGTGTGACCCAGGATGGGATGGG 251154
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 251155 AACTTAACCTTAATGCTCTACAAATGCAATGACCAACGATAGAAATTAATACCGAGAA 251214
QY 121 ArgPheValGlnLeuAlaPheLeuArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
DB 251215 AGGTTTGTGCAGATTGCTTTTATAGGCTAGAGGATCCACCAAGGAATCCCTATAGTGA 251274
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLeu 156
DB 251275 AACTATCAGGAAGCACAGATTGGTATTTTCCAAAGAGAAGAAACTC 251322

RESULT 7

BD430793_17

WPCOMMENT

Sequence split into 18 fragments LOCUS BD430793 Accession BD430793

Fragment Name Begin End

BD430793_00 1 110000
BD430793_01 100001 210000
BD430793_02 200001 310000
BD430793_03 300001 410000
BD430793_04 400001 510000
BD430793_05 500001 610000
BD430793_06 600001 710000
BD430793_07 700001 810000
BD430793_08 800001 910000
BD430793_09 900001 1010000
BD430793_10 1000001 1110000
BD430793_11 1100001 1210000
BD430793_12 1200001 1310000
BD430793_13 1300001 1410000
BD430793_14 1400001 1510000

BD430793_15 1500001 1610000
BD430793_16 1600001 1710000
BD430793_17 1700001 1765118

Continuation (18 of 18) of BD430793 from base 1700001 (BD430793 GENOME SEQUENCE AND POLI

Alignment Scores:

Pred. No.: 9 91e-67 Length: 65118
Score: 703.00 Matches: 131
Percent Similarity: 94.12% Conservatives: 13
Best Local Similarity: 85.62% Mismatches: 9
Query Match: 87.22% Indels: 0
DB: 6 Gaps: 0

US-08-957-709A-71 (1-156) x BD430793_17 (1-65118)

QY 1 MetLeuLeuProAspTrpLysLeuArgLysGluLeuLeuGluProPheSerGluGlu 20
DB 62045 ATGCTCTTCAGACTGGAAATAAGGAAAGAGATTTTAAATAAAGCCATTTCTCAGAGGAA 62104
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 62105 TCGCTTCAACAGCTGGCTAGACTTAAGGGTAGCAAGGAGCTTACATTCAGGGAAG 62164
QY 41 LeuLeuAspValGluLysGluGlyValValLeuProArgGluTyrAlaLeuLeu 60
DB 62165 TTCATAGATGTTGAAAGGAGGCAAGCTCATATACCTCCAAAGAAATATGCCCTAATA 62224
QY 61 LeuThrLeuGluArgLysLeuProAspValMetGlyAspMetLysLeuArgSer 80
DB 62225 CTGACCCCTAGAGAGATAAAGCTTCCAGATGATATTATGGGAGACATGAAGATAAGGAGC 62284
QY 81 SerLeuAlaArgGluGlyValLeuGlySerPheAlaTrpValAspProGlyTrpAspGly 100
DB 62285 AGCTAGCTAGGGAAGGTGTTCTAGCTCCTTTGCTGGTAGACCCGGGATGGACGGT 62344
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 62345 AATCTAACCTTGATGCTTTACATGCATCGAAAGGAGGTAAATTTTAAGGTACAAAGAG 62404
QY 121 ArgPheValGlnLeuAlaPheLeuArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
DB 62405 AGGTTTGTCCAGATAGCTTCTTAAGGCTTCAGGCTCCCGCAAGAAATCCATACAGGGC 62464
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArg 153
DB 62465 AACTATCAAGGAAGCAGGAGAATAGTCTCTCAAAAGAGA 62503

RESULT 8

CNSPAX06

LOCUS

Pyrococcus abyssi complete genome; segment 6/6.

Accession

AJ248288.1 GI:5458960

Version

complete genome.

Keywords

Pyrococcus abyssi

Pyrococcus abyssi

Archaea; Euryarchaeota; Thermococci; Thermococcales;

Thermococcaceae; Pyrococcus.

1 (bases 1 to 265118)

Gaspin, C., Cavaille, J., Erauso, G. and Bachelier, J.P.

Archaeal homologs of eukaryotic methylation guide small nucleolar

RNA: lessons from the Pyrococcus genomes

J. Mol. Biol. 297 (4), 895-906 (2000)

20202371

10736225

REFERENCE

AUTHORS

Lecompte, O., Ripp, R., Puzos-Barbe, V., Duprat, S., Heilig, R.,

Districh, J., Thierry, J.C. and Poch, O.

Genome evolution at the genus level: comparison of three complete

Genomes of hyperthermophilic archaea

Genome Res. 11 (6), 981-993 (2001)

21275479

PUBMED

11381026

REFERENCE 3
AUTHORS Cohen,G., Barbe,V., Flament,D., Galperin,M., Heilig,R., Ripp,R.,
Lecompte,O., Prieur,D., Poch,O., Quelleron,J., Thierry,J.C., Van
der Oost,J., Weissenbach,J., Zivanovic,Y. and Forterre,P.
TITLE An integrated analysis of the genome of the hyperthermophilic
archaeon Pyrococcus abyssi
JOURNAL Mol. Microbiol. 47 (6), 1495-1512 (2003)
MEDLINE 22511545
PUBMED 12622808
REFERENCE 4 (bases 1 to 265118)
Genoscope.
Direct Submission
TITLE Direct Submission
JOURNAL Submitted (06-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- web : www.genoscope.cns.fr)
COMMENT Join(AJ248283.1:1..307120,AJ248284.1:51..293250,
AJ248285.1:51..307120,
AJ248286.1:51..288440,AJ248287.1:51..304290,AJ248288.1:51..265118).
FEATURES
source Location/Qualifiers
1..265118
/organism="Pyrococcus abyssi"
/mol_type="genomic DNA"
/strain="Orsay"
/db_xref="taxon:29292"
164..1615
/genes="bgal"
164..1615
/genes="bgal"
/notes="PAB2376"
/codon_start=1
/transl_table=11
/product="bgaL-2 beta-galactosidase (EC 3.2.1.23)
(lactase)"
/protein_id="CAB50447.1"
/db_xref="GI:5458961"
/db_xref="GOA:Q9UYG5"
/db_xref="UniProt/TREMBL:Q9UYG5"
/translation="MEFYGVQVSQAFQFENGDPYRNLDTRSDWHWRDPPYNIKREL
VSGDPEEGINNYEYIDHRLAKELGLNAYQLTIEWSRIFPCPTYSVEVERDYG
FYSKIRKHELEKDELAKREVRHYENLVKLNKLFNFTFVTLNHNQTNFWLHDP
EVRVNTKEARAGWDERAIVPESKAAVVAWKFYVDFWATFDEPMTYALGYLAP
YVGPGLINPKRAKSVIMHOIITAHARAVDSIKFSDKPVGTILNIIPAYLNPRDSK
HVAENAPLDTFHNRLFEAVNGKLLDIDINGEYKVPKPLKRDNDGNNTYTRVRYKV
EPYKELPIVTFVGVGYSNAPNSISPDNNPTSDFGWEVYPKGLYDSTAAREYSE
MVFITENGADSKILRPRIYDVHVKVEKRLRENGINVLGYFHWALTNDYEWAMGPKI
RFLGYEVDPIITKERIPRRKSVEYIKKVKKEGIE"
1834..2283
/note="PAB1019"
/codon_start=1
/transl_table=11
/product="Predicted nucleic acid-binding protein, contains
PIN domain"
/protein_id="CAB50448.1"
/db_xref="GI:5458962"
/db_xref="UniProt/TREMBL:Q9UYG4"
/translation="MPLPDITPDSVTLKMHSEKNRLLRLLITLAKFTVYLSMLTVQR
YLTARKNVELEVLKDIYNIIVPVSSEIILRASKIEANLRKGVLDLEDIITA
YTAIEGSLITDEPKKYPPIRRFGDTPDLKFLREVLWVEKEII"
complement(2276..3079)
/note="PAB1341"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="CAB50449.1"
/db_xref="GI:5458963"
/db_xref="UniProt/TREMBL:Q9UYG3"
/translation="MPENLESFEAPKLLLEVLPLEDLPDKVKEKVICIEPHDPDC
VIMGGTIKLTERGEIVYACMTDGYMTGLDSSLTGHELATIRRESEESKLLGVK
KIYLNRTLPNSVDVGLRIIRKEKPDGVFLPDLPYEARHPDRHNTGFLAD
AVAFSLPNFSNVSDVEIGLPHQVSFIALLYTNKPNYFVDITDVMELKKAIRTHKSK
PDDWWEVNEPFLRTVALPYGKAGVYAGFRIMPLGLFYHITFPFADLI"
3192..4340
/note="PAB1020"

/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="CAB50450.1"
/db_xref="GI:5458964"
/db_xref="GOA:Q9UYG2"
/db_xref="UniProt/TREMBL:Q9UYG2"
/translation="MKEVSSVYKEILVKGLTDETRIELEMGKGIIEDEPDGIRYVR
FKDSAGKRRGTVIDEYVIFGFHRIKRIINLRSGIRIFKRGFEYVEBKVDGYNVR
VVMYKGMIGITRGGIFCPFTTERIPDFVQPEFFKDNPNLILVGENAGPSPYLVEGP
PVVKEDIQFLFDVQIKTRSLPVEERLKAIENHVEVFGKTKDDVDLYQII
ERLSKEGREGIIMKSPDMKKIKVYVTPYANINDIKIGARVYELPCGYTSRLAF
YLAEKRIKEEPEERVAKEIGSALLQPFVESIPVEQEEDEHLEFKVRVRIETAYKV
THPEKGLKLEIYDIBEIKDGRWITFKRLYPDATNEIRELIGKAFVD"
4351..5658
/note="PAB1021"
/codon_start=1
/transl_table=11
/product="Putative hydroxypyruvate reductase"
/protein_id="CAB50451.1"
/db_xref="GI:5458965"
/db_xref="UniProt/TREMBL:Q9UYG1"
/translation="MEDIALRLVKAIRAADPYKAVINSVRVEGNKLLVKDKPEIKG
KYVLAFGKACDMARJAEILDVEDGVATVKYGYKELKKIRVIEAGHPIDENSTR
GAIEGLKLEKVEEDIVFLISGGSSALFELPEDGISLEDLKLTNELLKSKAKIHE
INTVRKHISKVGKGLAKIKGTGIVLIIISDVVDNLEAIASGPTVKDPTFQDAKRI
LELYGIWDRVPESVRSHEKIGIRGEAETLKEELPNHNLFIASNISQCAVVEATK
LGFKAYMTTTLLEGAKDAGLFGISVQSIATKNPFPFVVLVFGGETTVTTEGEG
KGGPNOELALSATRKIKGLNAILVAFDTGDTGPTDAAGGIVDGETVEKLRRKGIDIE
KVLKEHNSYEALKKVGSLFTGPTGTNVSIIAIIITSQGR1"
complement(5633..8655)
/note="PAB1338"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="CAB50452.1"
/db_xref="GI:5458966"
/db_xref="UniProt/TREMBL:Q9UYG0"
/translation="MPYIYEARVKVGISSWVTSKVSWRDALAQLESRIYVVKYLRMGE
EKGESFPFSALTIGIRIPDELKNPKDKHFGLFYVPGGELLVIPTTEENLLIYD
VEKGFABFKVFEVLINGKTMLOLARIVGGAIMNESKWEKGLRWIKVKSARTQTE
RSVVVITDKRPVSPFSDLEDIEIEVEVMNGKVRWKARHGHINOSVSYLYIPDKK
VQLFVURLYLLTPPSAMEFIIKIADDFPLKSEFQSELMKELKELSLDEMEKQILVA
LYSGINPLSHQFLTEKIEIIEIYDMIDKGLLKIWMIRKIVDLTNEGRKIVNKLK
YGLVSN"
complement(6670..7425)
/note="PAB1337"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="CAB50453.1"
/db_xref="GI:5458967"
/db_xref="UniProt/TREMBL:Q9UYF9"
/translation="MDVGTLLIIVVLIGSIVSAWISVSIQQMKWQALMISRL
EVETKMKRGTGRGSEAKBEVAKQEVDRVSEEDIVRKIKQIIDKVKHNVANEARK
KBERLMLDVARGVALGFVSDENAFMRVLSEDFKRLMLAAYPPPPDRLNQ
MUTHVARTKLPTEVREKKGKGVKLSQDEALIRITSSITKSAISILDDLKRRGNP
AVTPLEIKLSQLEKCKKVEDLEKRLQLEAL"
complement(7468..9687)
/note="PAB1336"
/codon_start=1
/transl_table=11
/product="Methyl accepting chemotaxis protein"
/protein_id="CAB50454.1"
/db_xref="GI:5458968"
/db_xref="GOA:Q9UYF8"
/db_xref="UniProt/TREMBL:Q9UYF8"
/translation="MQFRKKIILMAIMSAVITMAIFGGVTIYFNFVYVGIKEIQAP
AIEQGLQALSAAEVAEKDFDFFAVDYGTMAVKEAVKAGKDSNETAREFL
SEFKIKELNKDVSYYVFGDEBGMVYPPDELPEGYDPRVPMYVQAKGKGAIIIE
PYRDAATGQWIIYAVPVVNGQFKVIGLDVFSVTLNETKKIKLGKCYIAIIDQN
GMLVHPNEEFINKLNIPOVPELKDLADELKKGKDEGVLYITFQGIKKIGGYKRLKT
GMIVVATVPLSDITGTATAINMKMVIQGFAPVGVFIIVAILWIIYKIASNTL

isolated PEF proteins, and methods for purifying and identifying same

JOURNAL Patent: US 6379553-A 42 30-APR-2002;

FEATURES Location/Qualifiers

source

1..740

/organism="unknown"

ORIGIN /mol_type="unassigned DNA"

Alignment Scores: Length: 740
Pred. No.: 4.07e-51 Matches: 104
Score: 543.00 Conserv: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 67.37% Gaps: 0
DB:

US-08-957-709A-71 (1-156) x AR207520 (1-740)

Qy 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspVal 72
Db 2 CCTCCAGGGAATACGCCTTAATCTAACCCTCGAGAGGATTAAGTTGCCCGACGATGTT 61
Qy 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
Db 62 ATGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAAGGGTTATTGGTCTTTTGGT 121
Qy 93 TrpValAspProGlyTyrAlaLeuIleLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
Db 122 TGGGTTGACCCAGGATGGGATGGAACTTAACACTAATGCTCTACATGCTCAAAATGAA 181
Qy 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
Db 182 CCTGTCGAATTAAGATATGGAGAGATTTGTGCAGATCGCATTTATTAAGGCTAGAGG 241
Qy 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
Db 242 CCGCAAGAAACCTTACAGAGGAACCTATCAGGGGAGCACAAGGTTAGCGTTTCAAAG 301
Qy 153 ArgLysLysLeu 156
Db 302 AGAAGAAACTC 313

RESULT 13
AR225643 LOCUS AR225643 740 bp DNA linear PAT 20-DEC-2002

DEFINITION Sequence 42 from patent US 644428.

ACCESSION AR225643

VERSION AR225643.1 GI:27263675

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 740)

AUTHORS Hogrefe, H.

TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes,

isolated PEF proteins, and methods for purifying and identifying

same

JOURNAL Patent: US 644428-A 42 03-SEP-2002;

FEATURES Location/Qualifiers

source

1..740

/organism="unknown"

ORIGIN /mol_type="genomic DNA"

Alignment Scores: Length: 740
Pred. No.: 4.07e-51 Matches: 104
Score: 543.00 Conserv: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 67.37% Gaps: 0
DB:

US-08-957-709A-71 (1-156) x AR225643 (1-740)

Qy 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspVal 72
Db 2 CCTCCAGGGAATACGCCTTAATCTAACCCTCGAGAGGATAAAGTTGCCCGACGATGTT 61
Qy 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
Db 62 ATGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAAGGGTTATTGGTCTTTTGGT 121
Qy 93 TrpValAspProGlyTyrAlaLeuIleLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
Db 122 TGGGTTGACCCAGGATGGGATGGAACTTAACACTAATGCTCTACATGCTCAAAATGAA 181
Qy 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
Db 182 CCTGTCGAATTAAGATATGGAGAGATTTGTGCAGATCGCATTTATTAAGGCTAGAGG 241
Qy 133 ProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
Db 242 CCGCAAGAAACCTTACAGAGGAACCTATCAGGGGAGCACAAGGTTAGCGTTTCAAAG 301
Qy 153 ArgLysLysLeu 156
Db 302 AGAAGAAACTC 313

RESULT 14

AR535060 LOCUS AR535060 740 bp DNA linear PAT 08-OCT-2004

DEFINITION Sequence 42 from patent US 6734293.

ACCESSION AR535060

VERSION AR535060.1 GI:53925880

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 740)

AUTHORS Hogrefe, H. and Hansen, C.J.

TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes,

isolated PEF protein, and methods for purifying and identifying

Patent: US 6734293-A 42 11-MAY-2004;

FEATURES Location/Qualifiers

source

1..740

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores: Length: 740
Pred. No.: 4.07e-51 Matches: 104
Score: 543.00 Conserv: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 67.37% Gaps: 0
DB:

US-08-957-709A-71 (1-156) x AR535060 (1-740)

Qy 53 ProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspVal 72
Db 2 CCTCCAGGGAATACGCCTTAATCTAACCCTCGAGAGGATAAAGTTGCCCGACGATGTT 61
Qy 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
Db 62 ATGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAAGGGTTATTGGTCTTTTGGT 121
Qy 93 TrpValAspProGlyTyrAlaLeuIleLeuThrLeuMetLeuTyrAsnAlaSerAsnGlu 112
Db 122 TGGGTTGACCCAGGATGGGATGGAACTTAACACTAATGCTCTACATGCTCAAAATGAA 181
Qy 113 ProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGly 132
Db 182 CCTGTCGAATTAAGATATGGAGAGATTTGTGCAGATCGCATTTATTAAGGCTAGAGG 241

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:26:04 ; Search time 30.9767 Seconds
(without alignments)
1942.879 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPDWKIRKELIIPFSEE.....PYRGNYQSGTRLAFSKKKL 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1722976 seqs, 385795295 residues

Total number of hits satisfying chosen parameters: 1722976

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA.*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
 - 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	218	27.0	189	9	US-09-738-626-6640
2	210	26.1	191	14	Sequence 6640, Ap
3	152	18.9	195	16	Sequence 11598, A
4	138	17.1	190	15	Sequence 116832
5	134.5	16.7	177	10	Sequence 412, App
6	121	15.0	147	17	Sequence 247, App
7	121	15.0	181	18	Sequence 4752, App
8	121	13.9	158	10	Sequence 3527, Ap
9	111	13.8	164	15	Sequence 239, App
10	111	13.8	164	15	Sequence 32, Appl
11	111	13.8	229	15	Sequence 409, App
					Sequence 10, Appl
					Sequence 11998, A
					Sequence 116832
					Sequence 412, App
					Sequence 247, App
					Sequence 4752, App
					Sequence 3527, Ap
					Sequence 239, App
					Sequence 32, Appl
					Sequence 409, App
					Sequence 10, Appl

Query Match 27.0%; Score 218; DB 9; Length 189;

ALIGNMENTS

RESULT 1

US-09-738-626-6640
; Sequence 6640, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6640
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6640

12	111	13.8	229	15	US-10-336-472-30	Sequence 30, Appl
13	111	13.8	252	15	US-10-136-728-8	Sequence 8, Appl
14	111	13.8	252	15	US-10-336-472-28	Sequence 28, Appl
15	111	13.8	252	16	US-10-408-765A-2116	Sequence 2116, Ap
16	102	12.7	178	9	US-09-970-711-2	Sequence 2, Appl
17	99.5	12.3	254	10	US-09-975-719-241	Sequence 241, App
18	97	12.0	148	16	US-10-627-592-10	Sequence 10, Appl
19	97	12.0	150	16	US-10-627-592-8	Sequence 8, Appl
20	93	11.5	171	16	US-10-437-963-173883	Sequence 173883, Ap
21	90.5	11.2	171	16	US-10-425-115-256676	Sequence 256676, Ap
22	90	11.2	222	16	US-10-739-930-9568	Sequence 9568, Ap
23	89.5	11.1	175	15	US-10-424-599-269427	Sequence 269427, Ap
24	87.5	10.9	1150	9	US-09-946-239-9	Sequence 9, Appl
25	87.5	10.9	1259	15	US-10-282-122A-56089	Sequence 56089, A
26	84	10.4	701	15	US-10-282-122A-55632	Sequence 55632, A
27	82	10.2	863	15	US-10-369-493-3655	Sequence 3655, Ap
28	82	10.2	884	17	US-10-732-923-23250	Sequence 23250, A
29	81.5	10.1	902	14	US-10-231-778-222	Sequence 222, App
30	81	10.0	154	9	US-09-712-363-250	Sequence 250, App
31	80	9.9	149	9	US-09-738-626-5594	Sequence 5594, Ap
32	79.5	9.9	502	14	US-10-006-852-18	Sequence 18, Appl
33	79.5	9.9	502	17	US-10-732-923-6235	Sequence 6235, Ap
34	79	9.8	267	15	US-10-369-493-20803	Sequence 20803, A
35	78.5	9.7	457	15	US-10-369-493-11845	Sequence 11845, A
36	78	9.7	201	10	US-09-988-067B-50	Sequence 50, Appl
37	78	9.7	290	16	US-10-437-963-120944	Sequence 120944, Ap
38	76.5	9.5	146	15	US-10-335-977-7336	Sequence 7336, Ap
39	76.5	9.5	154	15	US-10-335-977-7337	Sequence 7337, Ap
40	76.5	9.5	532	16	US-10-437-963-115812	Sequence 115812, A
41	76	9.4	402	15	US-10-282-122A-53348	Sequence 53348, A
42	75	9.3	206	15	US-10-335-977-8955	Sequence 8955, Ap
43	75	9.3	209	15	US-10-335-977-8956	Sequence 8956, Ap
44	75	9.3	815	15	US-10-282-122A-52922	Sequence 52922, A
45	74.5	9.2	499	15	US-10-424-599-248456	Sequence 248456, A

```
Best Local Similarity 33.7%; Pred. No. 1.6e-16;
Matches 57; Conservative 28; Mismatches 60; Indels 24; Gaps 4;
QY 1 MLLPDKIRK-----EIIIEPSESLOPAGYDLVRGR--EAFVKGK-----40
Db 1 MLLSDRDIRKSIDAGDLGIEPDABELIQPSSVDVMDRYFRVFNNSKYTHIDPKLNODEL 60
QY 41 --LIDVEGKVIIPREYALITLIERIKLPDVMGDMKIRSSLAREGVI--GSPAWDP 96
Db 61 TSLVEVDGEGFVLHPGEFVLASTLEKFTLPAHLAAGLEKSSGLRGLLTHSTAGFIDP 120
QY 97 GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYQGS 145
Db 121 GFSGYITLSELSNVANLPTLWFGMKVQQLAFQMSPAETPYGSGKLG 169
```

```
RESULT 2
US-10-156-761-11998
; Sequence 11998, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11998
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11998
```

```
Query Match 26.1%; Score 210; DB 14; Length 191;
Best Local Similarity 31.4%; Pred. No. 1.3e-15;
Matches 53; Conservative 32; Mismatches 60; Indels 24; Gaps 3;
QY 1 MLLPDKIRKEI-----LIEPFSESLOPAGYDLVRGR-----EAFVK 38
Db 1 VLLSKDINAEIDAGRVIDPDYDESMVQSSIDVRLDRYFRVFNHRYPHIDPSVEQADL 60
QY 39 GKLIDVEGKVIIPREYALITLIERIKLPDVMGDMKIRSSLAREGVI--GSPAWDP 96
Db 61 TRLVEPGEDEPFIHPGEFVLASTVEVLSLDDLASRLGKSSGLRGLVTHSTAGFIDP 120
QY 97 GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYQGS 145
Db 121 GFSGHVTLSELSNLATLPIKLMGKIGQLCMFRLSPAEFPYGSRDYGS 169
```

```
RESULT 3
US-10-437-963-116832
; Sequence 116832, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
```

```
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 116832
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(195)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_20296C.1.pep
US-10-437-963-116832
```

```
Query Match 18.9%; Score 152; DB 16; Length 195;
Best Local Similarity 35.0%; Pred. No. 6.5e-09;
Matches 42; Conservative 18; Mismatches 56; Indels 4; Gaps 3;
QY 22 LQPAGYDLRVGREAFVVGKLIIDVEKEGKVITPPREYALITLIERIKLPDVMGDMKIRSS 81
Db 54 LSGPGEQVSAQLEKIMSDEIILAEGEA--FFLHPGELALATTLESVKLPDNIWGWLDRSS 112
QY 82 LAREGVIGSFA--WVDP--GWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPY 138
Db 113 LARLGLMVHTAHRIDPXGWHGKIVLEFPFNAGKULPAURLNNAIGALSFEVLGAAAKPY 172
```

```
RESULT 4
US-10-289-762-412
; Sequence 412, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 412
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-289-762-412
```

```
Query Match 17.1%; Score 138; DB 15; Length 190;
Best Local Similarity 24.4%; Pred. No. 2.6e-07;
Matches 40; Conservative 31; Mismatches 59; Indels 34; Gaps 5;
QY 13 LIEPP-----SBES-----LQPAGYDLRVGRE-----AFVKGKLI 42
Db 18 MIHPVNGQVNVNVEETGEKLISSYGYDLRLSREFKVFTNVNYSVVDPKCFTEIDFI 77
QY 43 DVEKEGKVIIPREYALITLIERIKLPDVMGDMKIRSSLAREGVIGSFAWVDPGWDNL 102
Db 78 SI-TDDVCIVFPNSFALARSVEYFRIPRNVLTMCIGKSTYARCGIIVNVTPPEPEWEGHV 136
QY 103 TMLYNASNEPVELRYGERFVQIAFIRLEGPARNPY---RGNVQ 143
Db 137 TIBISNTTLPKAIYANEGIAQVLFFSSTTCSEVSADRKGYIQ 180
```

```
RESULT 5
US-09-975-719-247
; Sequence 247, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
```

```

; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 247
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-247

```

Query Match	16.7%	Score 134.5;	DB 10;	Length 177;
Best Local Similarity	26.5%;	Pred. No. 5.9e-07;		
Matches	41;	Conservative 31;	Mismatches 64;	Indels 19; Gaps 4;
Qy	8	IRKEILLIEPSESRL---	QPAGYDLRVG-----	REAFVK---GKLDIVKEGK 49
Db	12	VRDGKLLKLAHRELTT	PEGVGFDLRAGLSRLTV	GGSLRESRRTRTPASEWILADPDC 71
Qy	50	WVTPPREYALILTLER	IKLDDVNGMKIRSSLA	REGVIGSFAMVDPGWDGNTLMLYNA 109
Db	72	FVMEPGKTYLASTWEE	FDLEDLALAFPPSTL	FRSGITFSSSVLPFGYVGPMTFALTNN 131
Qy	110	SNPEVLRYGERFVQIA	IFIRLEGPARNPYRG	NYOG 144
Db	132	HSASAFETQIGARFAH	VIFQAVSGDI -GRYK	GOWOG 165

```

RESULT 6
US-10-472-928-4752
; Sequence 4752, Application US/10472928
; Publication No. US2005020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: Seqwin99, version 1.03
; SEQ ID NO 4752
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: deoxyuridine 5'triphosphate nucleotidohydrolase, putative
; OTHER INFORMATION: Cellular location: cytoplasm
; OTHER INFORMATION: Similar to strain R6 sequence 15902067 (2.E-81)
US-10-472-928-4752

```

Query Match	15.0%;	Score 121;	DB 17;	Length 147;
Best Local Similarity	28.1%;	Pred. No. 1.6e-05;		
Matches	45;	Conservative 23;	Mismatches 56;	Indels 36; Gaps 5;
QY	7	KIRKETLIEPSESLOP-----	ACYDLRVG-REAFVKGKGLIDVKEGKVVIPPEYA	58
Db	2	KIRGFELVSSFDENLLPKRETAHAACYDLKAVRTVAFGEIVLVPTGKAYMQPTEVL	61	
QY	59	LILTLERIKLPDDVMGDMKIRSSLARE-----	GVIGSFAMVDPGWDGNLTMLYNA	109
Db	62	YLVD-----	RSSNPRKKGLVLNSVGVIDGDYGNFGNEGHPAQMKNI	105
QY	110	SNPEVLRYGERFVQ---IAFIRLEGPARNPYRGNYOGST	146	

[illegible]

```

RESULT 8
US-09-975-719-239
; Sequence 239, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975.719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 239
; LENGTH: 181
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-239

Query Match 13.9%; Score 112; DB 10; Length 181;
Best Local Similarity 24.9%; Pred. No. 0.00024;
Matches 43; Conservative 28; Mismatches 76; Indels 26; Gaps 5

Qy 2 LIPDMKIRKEILIEPFS-----EESLPQAGYDLRVGREA-----FVKGKLIIDVEK 46
Db 13 LVAQCAICEKTLDFEBSLKIQIOLGHTYYEPKSLPASASIVYGSHPAFPSTFTFLEPK----EI 68

Qy 47 EGVKVTIPPEYALILTLERIKLPDDVMGDMKIRSLAR--EGVIGSPAWVDPCMGDGNLTL 104
Db 69 QONVLKSGEQVITCKIRYKIPLDYFLGLVOTKGLTLARLFVQVTCNDGQVEPGFDGYVTL 128

Qy 105 MLYNASNEPVELRYGERFVQIAFTIRLEGAPARNRYGNY-----QGSTRLAFSK 152
Db 129 EIVNMPWTIEIPAVSDIAQLVYVKCSTASPEYHGRYMDAAKKGPTLAVFRK 181

RESULT 9
US-10-336-472-32
; Sequence 32, Application US/10336472
; Publication No. US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Hallinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.

```



```
; FILE REFERENCE: BIOBANK.012A
; CURRENT APPLICATION NUMBER: US/10/733,878
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 60/432699
; PRIOR FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 60/485027
; PRIOR FILING DATE: 2003-07-03
; NUMBER OF SEQ ID NOS: 535
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 409
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-733-878-409

Query Match      13.8%; Score 111; DB 16; Length 164;
Best Local Similarity 30.7%; Pred. No. 0.00027;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

QY 21 SIQAGYDLRVGREAFVKGKLDVKEGKVVPVPPREYALILTLERIKLPDDVGMGMKIRS 80
Db 42 SARAAGYDLYSAYD-----YTIPPMEKAVVKTQIQLPSPGCGYGRVAPRS 86

QY 81 SIAREGVIGSFAMV-DPGWGDGNTLMLYNASNEPVELRYGERFVQIAFIRLEGP 133
Db 87 GLAAKHFDVVGAGVIDEDYRGNGVGVLFNFGKEKEFKGDRIAQLICERIFYP 140

RESULT 11
US-10-136-728-10
; Sequence 10, Application US/10136728
; Publication No. US20030236188A1
; GENERAL INFORMATION:
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Li, Li
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Stone, David J.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Anderson, David W.
; APPLICANT: Patturajan, Meera
; APPLICANT: Gerlach, Valerie L.
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Smithson, Glenda
; APPLICANT: MacDougall, John R.
; APPLICANT: Mezes, Peter S.
; APPLICANT: Perman, John A.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: No. US20030236188A1el Human Proteins, Polynucleotides Encoding Th
; FILE REFERENCE: 21402-347 D (Cura 647 Other)
; CURRENT APPLICATION NUMBER: US/10/136,728
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 60/288,395
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/289,087
; PRIOR FILING DATE: 2001-05-07
; PRIOR APPLICATION NUMBER: 60/289,619
; PRIOR FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: 60/289,818
; PRIOR FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: 60/289,817
; PRIOR FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: 60/290,194
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/290,753
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/291,189
; PRIOR FILING DATE: 2001-05-15

; FILE REFERENCE: BIOBANK.012A
; CURRENT APPLICATION NUMBER: US/10/733,878
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 60/432699
; PRIOR FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 60/485027
; PRIOR FILING DATE: 2003-07-03
; NUMBER OF SEQ ID NOS: 535
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 409
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-733-878-409

Query Match      13.8%; Score 111; DB 16; Length 164;
Best Local Similarity 30.7%; Pred. No. 0.00027;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

QY 21 SIQAGYDLRVGREAFVKGKLDVKEGKVVPVPPREYALILTLERIKLPDDVGMGMKIRS 80
Db 42 SARAAGYDLYSAYD-----YTIPPMEKAVVKTQIQLPSPGCGYGRVAPRS 86

QY 81 SIAREGVIGSFAMV-DPGWGDGNTLMLYNASNEPVELRYGERFVQIAFIRLEGP 133
Db 87 GLAAKHFDVVGAGVIDEDYRGNGVGVLFNFGKEKEFKGDRIAQLICERIFYP 140

RESULT 12
US-10-336-472-30
; Sequence 30, Application US/10336472
; Publication No. US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Spaderma, Steven K.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Ort, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
```

```
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,681
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 30
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-136-472-30

Query Match 13.8%; Score 111; DB 15; Length 229;
Best Local Similarity 30.7%; Pred. No. 0.00042;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

QY 21 SLQAGYDLVRGREAFAVKGKLDVKEGKVIPPPEYALILTLEIKLPDDVGMGMKIRS 80
DB 107 SARAAGYDLYSAYD-----YTIPPEKAVVKTDIQIALPSCGYGRVAPRS 151

QY 81 SLAREGVITGSPAWV-DPGWDGNLTMLYNASNEPVLELYRGRFVQIAIRLEGP 133
DB 152 GLAAKHFDVGVAGVIDEDYRGNVGVLFNFGKEFEVKKGDRIAQLICERIFYP 205

RESULT 13
US-10-136-728-8
; Sequence 8, Application US/10136728
; Publication No. US20030236188A1
; GENERAL INFORMATION:
; APPLICANT: Spytex, Kimberly A.
; APPLICANT: Li, Li
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Stone, David J.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Anderson, David W.
; APPLICANT: Patturajan, Meera
; APPLICANT: Gerlach, Valerie L.
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gorman, Linda
; APPLICANT: Zethusen, Bryan D.
; APPLICANT: Smithson, Glenda
; APPLICANT: MacDougall, John R.
; APPLICANT: Mezes, Peter S.
; APPLICANT: Perman, John A.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: No. US20030236188A1el Human Proteins, Polynucleotides Encoding TH
; FILE REFERENCE: 21402-347 D (Cura 647 Other)
; CURRENT APPLICATION NUMBER: US/10/136,728
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 60/288,395
```

```
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/289,087
; PRIOR FILING DATE: 2001-05-07
; PRIOR APPLICATION NUMBER: 60/289,619
; PRIOR FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: 60/289,818
; PRIOR FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: 60/289,817
; PRIOR FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: 60/290,194
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/290,753
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/291,189
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/292,374
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/293,107
; PRIOR FILING DATE: 2001-05-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 132
; SEQ ID NO 8
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-136-728-8

Query Match 13.8%; Score 111; DB 15; Length 252;
Best Local Similarity 30.7%; Pred. No. 0.00048;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

QY 21 SLQAGYDLVRGREAFAVKGKLDVKEGKVIPPPEYALILTLEIKLPDDVGMGMKIRS 80
DB 130 SARAAGYDLYSAYD-----YTIPPEKAVVKTDIQIALPSCGYGRVAPRS 174

QY 81 SLAREGVITGSPAWV-DPGWDGNLTMLYNASNEPVLELYRGRFVQIAIRLEGP 133
DB 175 GLAAKHFDVGVAGVIDEDYRGNVGVLFNFGKEFEVKKGDRIAQLICERIFYP 228

RESULT 14
US-10-136-472-28
; Sequence 28, Application US/10336472
; Publication No. US20040043929A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Millet, Isabelle
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Spaderna, Steven K.
```

```
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Sheno, Suresh G.
; APPLICANT: Ort, Tatiana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,691
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 28
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-28

Query Match      13.8%; Score 111; DB 15; Length 252;
Best Local Similarity 30.7%; Pred. No. 0.00048;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

Qy 21 SLOPAGYDLRVGREAFVKGKLIIDVEKGVVIPPYVALITLTERIKLPDDVMGDMKIRS 80
Db 130 SARAAGYDLYSAYD-----YTIPPEKAVVKTDTQIALPSCGYGRVAPRS 174

Qy 81 SLAREGVIGSFANV-DFGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGP 133
Db 175 GLAAKHFDVAGVIDEDYRGNGVWLFNFGKKEFEVKGKGDRIAQLICERIFYP 228

RESULT 15
US-10-408-765A-2116
; Sequence 2116, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
```

```
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2116
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2116

Query Match      13.8%; Score 111; DB 16; Length 252;
Best Local Similarity 30.7%; Pred. No. 0.00048;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

Qy 21 SLOPAGYDLRVGREAFVKGKLIIDVEKGVVIPPYVALITLTERIKLPDDVMGDMKIRS 80
Db 130 SARAAGYDLYSAYD-----YTIPPEKAVVKTDTQIALPSCGYGRVAPRS 174

Qy 81 SLAREGVIGSFANV-DFGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGP 133
Db 175 GLAAKHFDVAGVIDEDYRGNGVWLFNFGKKEFEVKGKGDRIAQLICERIFYP 228

Search completed: July 2, 2005, 12:34:05
Job time : 31.9767 secs
```

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 11:06:33 ; Search time 32.9302 Seconds
(without alignments)
1832.195 Million cell updates/sec

```

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPDWKIRKEILIEPSEE.....PYRGNVOGSTRLAFSKRKKL 156

```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs. 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 200000000
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45

```
Database : A_Geneseqp_16Dec04:
1: Geneseqp1380rs:
2: Geneseqp1390rs:
3: Geneseqp2000rs:
4: Geneseqp2001rs:
5: Geneseqp2002rs:
6: Geneseqp2003as:
7: Geneseqp2003bs:
8: Geneseqp2004rs:
```

SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length	%			
1	806	100.0	156	2	AAW72847	Polymeras	
2	703	87.2	154	4	AAB96424	Putative	
3	242	30.0	173	4	AAAS1078	AAr51078	
4	227.5	28.2	198	8	ADK16665	Nanoarcha	
5	220.7	27.3	193	7	ADM25685	Hyperther	
6	218	27.0	189	4	AAG92886	C glutami	
7	196	24.3	160	4	AAB80139	Corynebac	
8	191.5	23.8	193	5	ABP66173	Bifidobac	
9	185.5	23.0	197	6	ADA33378	Acinetoba	
10	181	22.5	188	2	AAW89818	protein e	
11	181	22.5	190	2	AAW89897	Antigen 3	
12	172	21.3	202	7	ABO68916	Pseudomon	
13	170.5	21.2	194	6	ABM68738	Photorhab	
14	168.5	20.9	192	8	ADL05995	M. catarr	
15	168.5	20.9	204	7	ADF04778	Bacterial	
16	164	20.3	154	7	ADM26960	Hyperther	
17	164	20.3	217	7	ABO61782	Klebsiell	
18	156	19.4	188	6	ABP78401	N. gonorr	
19	155.5	19.3	139	5	ABU52178	Helicobac	
20	154.5	19.2	191	8	ADO57546	Actinobac	
21	149.5	18.5	190	2	AAV37187	Protein i	
22	143.5	17.8	1881	8	ADK15470	Taro bac	
23	138	17.1	190	2	AAV34994	Protein 1	
24	134.5	16.7	177	2	ABU29169	Amino aci	
25	121	15.0	147	6	ABU02796	S. pneumo	

ALIGNMENTS

RESULT 1

AAW72847
ID AAW72847 standard; protein; 156 AA.

AAW72847:

17-OCT-2003 (revised)

DT	17-OCT-2003	(revised)
DT	01-MAR-1999	(first entry)

DE Polymerase enhancing factor P45 (dUTPase) component.

AA Polymerase enhancing factor; PEF; dutPase; DNA polymerase; PCR;
KW amplification; sequencing; replication.
KW

OS *Pyrococcus furiosus*: strain DSM 3638.

XX	Key	Location/Qualifiers
FH		

FT	key	Proact
FT	Peptide	2. .15

FT peptide used to generate primers"

WO9842860-A1.

01-OCT-1998

XX
PF
20-MAR-1998: 98WO-US005497XX
PP 21-MAR-1997. 97HS-00822774

PR 21-MAR-1997; 97US-00822774.
PR 24-OCT-1997. 97US-00957709

XX
PA (STRA-) STRATAGENEXX
BT
Hogrefe U
Hansen C.T.

XX DB WDT. 1000-E4238A/AC

BR WPI; 1998-542284/
DP N-PSDB: AAV63860

PT Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
XX amplification, sequencing and replication.

XX
PS
Claim 17: Page 43: 161pp: English.

This is the amino acid sequence of the P45 component of the polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. The sequence is predicted from a DNA sequence (see AAV63860) obtained from genomic DNA by PCR. P45 and P50 (see AAW72844) are the predominant components of PEF, which acts to enhance the activity of *P. furiosus* DNA polymerase. P45 functions as a dUTPase, and can be used to enhance nucleic acid

CC replication, polymerisation or PCR reactions. The invention provides
CC novel extracts, proteins and complexes that improve the polymerisation
CC activity of nucleic acid polymerases, as well as DNA constructs and
CC antibodies. Also included are methods for identifying compositions with
CC PEF activity, for purifying and using these compositions, and specific
CC extracts, proteins and complexes that function to enhance polymerase
CC activity. Nucleic acid polymerase reactions can be enhanced (claimed) by
CC mixing a nucleic acid template, at least 1 polymerase and a composition
CC having polymerase enhancing activity. Kits are provided for replicating
CC nucleic acids, for site-directed mutagenesis, for nucleic acid sequencing
CC or amplification (preferably PCR or RT-PCR). (Updated on 17-OCT-2003 to
CC standardise OS field)
XX Sequence 156 AA;
SQ

Query Match 100.0%; Score 806; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. NO. 5.4e-89;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLLPDWKIRKELIIEPFSESLQAPGYDLRVGREAFVKGKLDVKEGKVIPPPEYALI 60
DB 1 MLLPDWKIRKELIIEPFSESLQAPGYDLRVGREAFVKGKLDVKEGKVIPPPEYALI 60
QY 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSFAPVDPGWDGNTLMLYNASNEPVELRYGE 120
DB 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSFAPVDPGWDGNTLMLYNASNEPVELRYGE 120
QY 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL 156
DB 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL 156

RESULT 2
AAB96424
ID AAB96424 standard; protein; 154 AA.
XX
AC AAB96424;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative P. abyssi deoxycytidine deaminase.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
PN FR2792651-A1.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99FR-00005034.
XX
PR 21-APR-1999; 99FR-00005034.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX WPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode proteins
PT useful in industry.
PT
XX
PS Claim 7; Page 1118; 1657pp; French.
XX
CC The present invention relates to the genomic sequence of Pyrococcus
CC abyssi (see AAF86431 and AA41223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to

CC 110 degrees centigrade. Note: This patent is in the same patent family as
CC WO200065062, which contains additional sequences as shown in AAB99132-
CC AAB99143, AAH75903-AAH75920 and AAG66436
XX
SQ Sequence 154 AA;
Query Match 87.2%; Score 703; DB 4; Length 154;
Best Local Similarity 85.6%; Pred. NO. 1.6e-76;
Matches 131; Conservative 13; Mismatches 9; Indels 0; Gaps 0;
QY 1 MLLPDWKIRKELIIEPFSESLQAPGYDLRVGREAFVKGKLDVKEGKVIPPPEYALI 60
DB 1 MLLPDWKIRKELIIEPFSESLQAPGYDLRVGREAFVKGKLDVKEGKVIPPPEYALI 60
QY 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSFAPVDPGWDGNTLMLYNASNEPVELRYGE 120
DB 61 LTLERIKLPDDVGMGDMKIRSSILAREGVIGSFAPVDPGWDGNTLMLYNASEKIVILRYKE 120
QY 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKR 153
DB 121 RFVQIAFIRLEAPAKNPYRGNYQGSRRIVLSKR 153
RESULT 3
AARS1078
ID AARS1078 standard; protein; 173 AA.
XX
AC AARS1078;
XX
DT 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 10-MAR-2003 (revised)
DT 27-MAY-1994 (first entry)
XX
DE Polypeptide encoded by pDam-L3 plasmid fragment.
XX
KW Ligase, thermostable; thermostability; Desulfohalobus ambivalens;
KW archaeobacteria; detection; ligation; ATP; adenosine triphosphate.
XX
OS Acidianus ambivalens.
XX
PN DE4217134-A1.
XX
PD 25-NOV-1993.
XX
PF 23-MAY-1992; 92DE-04217134.
XX
PR 23-MAY-1992; 92DE-04217134.
XX
PA (BOF) BOEHRINGER MANNHEIM GMBH.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Rueger R, Kessler C, Kaletta C, Jarsch M, Kletzin A;
XX
XX WPI; 1993-378402/48.
DR N-PSDB; AAQ62300.
XX
PT Thermostable ligase from archaeobacteria - and DNA coding for it, useful
PT for nucleic acid detection.
XX
PS Example 3; Fig 1; 26pp; German.
XX
CC A sequence encoding a thermostable DNA ligase was cloned into the plasmid
CC pDam-L3. The DNA ligase is ATP dependent and is useful for detecting
CC nucleic acids by hybridising two oligonucleotides with adjacent sequences
CC of the target nucleic acid, ligating the two oligonucleotides (using the
CC ligase) and detecting the ligation product. Analysis of the fragment
CC revealed three other open reading frames (ORF's). This is a polypeptide
CC encoded by ORF3 (Updated on 10-MAR-2003 to add missing OS field.)
CC (Updated on 25-MAR-2003 to correct PN field.) (Updated on 24-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 173 AA;

DE Acinetobacter baumannii protein #539.
 XX
 KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
 KW plant biocontrol agent.
 XX
 OS Acinetobacter baumannii.
 XX
 PN US6562958-B1.
 XX
 PD 13-MAY-2003.
 XX
 PF 04-JUN-1999; 99US-00328352.
 XX
 PR 09-JUN-1998; 98US-0088701P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Breton G, Bush D;
 XX
 XX WPI; 2003-576092/54.
 DR
 DR N-PSDB; ADA29252.
 XX
 XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
 PT for diagnosing a bacterial disease, as components of antibacterial
 PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
 PT plants.
 PS Example; SEQ ID NO 4665; 328pp; English.
 XX
 XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
 CC The A. baumannii nucleic acids and polypeptides are useful as reagents
 CC for diagnosing a bacterial disease, as components of antibacterial
 CC vaccines, as targets for antibacterial drugs, to detect the presence of
 CC A. baumannii and other Acinetobacter species in a sample, in screening
 CC compounds for the ability to interfere with the A. baumannii life cycle
 CC or to inhibit A. baumannii infection, and as biocontrol agents for
 CC plants. The present sequence represents the amino acid sequence of an A.
 CC baumannii protein.
 XX
 SQ Sequence 197 AA;
 Query Match 23.0%; Score 185.5; DB 6; Length 197;
 Best Local Similarity 27.4%; Pred. No. 1.2e-13;
 Matches 48; Conservative 31; Mismatches 63; Indels 33; Gaps 4;
 QY 10 KEILTEPSEESLQ-----PAGYDLRVGRE-----AFVKGK 40
 DB 23 KHGMIEPYAENQVRFDKNGEKLISYGVSSGYDVRCAREFKVTNNVHSAIVDPKNFDEKS 82
 QY 41 LIDVEKEGVIPPREYALILTLERIKLPDDVGMGDMKIRSSLSAREGVIGSPFVWDPGWDG 100
 DB 83 FIDIESD-VCIIPPNFSFALARTIEYFRIPRNVLTVCGLGKSTYARCGIIVNVTPLPEWEG 141
 QY 101 NUTLMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYR---GNYQGSTRLAFSK 152
 DB 142 HVTLEFSNTNLPARIYAGEVAQMLPFESDEVCSYKDRGKYQGQGTGVTLPK 196
 RESULT 10
 AAW89818
 ID AAW89818 standard; protein; 188 AA.
 XX
 AC AAW89818;
 XX
 XX 20-MAR-2003 (revised)
 DT 18-FEB-1999 (first entry)
 XX
 DE Protein encoded by clone B17 ORF4.
 XX
 KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
 XX
 OS Helicobacter pylori.

XX WO9849314-A2.
 XX
 PD 05-NOV-1998.
 XX
 PF 25-APR-1998; 98WO-US008487.
 XX
 PR 25-APR-1997; 97US-0045107P.
 PR 14-OCT-1997; 97US-0061958P.
 XX
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 XX
 PI Chow TP, Fry KE, Lim MY, Mcatee CP;
 XX
 DR WPI; 1999-009433/01.
 DR N-PSDB; AAV90555.
 XX
 XX New Helicobacter pylori antigens and related nucleic acid sequences -
 PT useful in serological diagnosis and protective vaccines, providing long-
 PT lasting immune response.
 XX
 PS Claim 15; Page 104-105; 402pp; English.
 XX
 XX The present sequence represents a Helicobacter pylori antigenic protein
 CC that is characterised by immunoreactivity with H. pylori-positive
 CC antisera. The proteins are highly immunogenic and induce a long-lasting
 CC immune response that persists even after antimicrobial treatment. In
 CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
 CC highly sensitive and specific. The specification also describes 69
 CC previously unrecognised immunogenic cluster families. H. pylori antigens
 CC are used to detect H. pylori-specific antibodies, for diagnosing
 CC infection or to confirm eradication of infection, and in vaccines to
 CC protect against H. pylori infection and related diseases (gastritis,
 CC peptic ulcer, gastric adenocarcinoma/lymphoma). (Updated on 20-MAR-2003
 CC to correct PF field.)
 XX
 SQ Sequence 188 AA;
 Query Match 22.5%; Score 181; DB 2; Length 188;
 Best Local Similarity 28.0%; Pred. No. 3.9e-13;
 Matches 47; Conservative 29; Mismatches 64; Indels 28; Gaps 4;
 QY 13 LIEPFSE-----SLQPAGYDLRVGRE-----AFVKGK-----LIDVEKE 47
 DB 18 MISPFCEKQVGNVTSYGLSSGYDIRVSEFMLFDKNKALIDPKNFPNNATKIDASK 77
 QY 48 GKVVIPPREYALILTLERIKLPDDVGMGDMKIRSSLSAREGVIGSPFVWDPGWDGNTLMY 107
 DB 78 GYFILPANAFALAHITIEYFKMPKDTLAIKLGKSTYARCGIIVNVTPEPEFEGYITIEIS 137
 QY 108 NASNEPVELRYGERFVQIAFIRLEGPARNPYR---GNYQGSTRLAFSK 152
 DB 138 NNTNLPKAVYANEGIAQVVFLOQDCEQSYKDRGKYQGQGTGVTLPK 185
 RESULT 11
 AAW89897
 ID AAW89897 standard; protein; 190 AA.
 XX
 AC AAW89897;
 XX
 XX 20-MAR-2003 (revised)
 DT 18-FEB-1999 (first entry)
 XX
 DE Antigen 3 from cluster 28c.
 XX
 KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9849314-A2.
 XX


```
PR 07-FEB-2001; 2001FR-00001659.
XX (INSP ) INST PASTEUR.
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Duchaud E, Taouit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
XX WPI; 2003-148459/14.
DR
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
XX Claim 2; SEQ ID NO 1835; 1205pp; French.
XX
XX The invention relates to the isolation of genes and their encoded
XX proteins from Photorhabdus luminescens. The isolated sequences are
XX sources of probes and primers for detecting the genome of P. luminescens
XX and related species; to study polymorphisms; for gene analysis and for
XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX polypeptides encoded by the genes are used for detection/identification
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX carry a gene-containing vector are used to select compounds that
XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX animals or microorganisms other than P. luminescens and are able to alter
XX response or sensitivity to toxins and antibiotics produced by P.
XX luminescens. Cells transformed to express the genes are useful for
XX recombinant production of the proteins, particularly toxins and
XX antibacterials useful as insecticides, bactericides and fungicides. The
XX genes, proteins, vectors containing the genes and Ab are also useful
XX therapeutically (to treat microbial infection by bacteria or fungi that
XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX biopesticides. Other uses of the genes and the proteins are as virulence
XX factors and for identifying targets of human diseases for which P.
XX luminescens is a model (particularly plague and whooping cough). This
XX sequence represents one of the isolated P. luminescens proteins
XX
XX Sequence 194 AA;
XX
XX Query Match 21.2%; Score 170.5; DB 6; Length 194;
XX Best Local Similarity 29.1%; Pred. No. 7.7e-12;
XX Matches 51; Conservative 28; Mismatches 61; Indels 35; Gaps 6;
XX
QY 6 WKIRKEILIEPSE-ESLQPAGYDLRVGRE-----AFVK-----GKLIDVEKE 47
DB 11 WLDEGKLVIAAPPIERINGATADVRLGNQFRVFCGHTAAAYIDLSPKDEVSAAALDRVMS 70
QY 48 GKWVIP-----PREVALILTLERIKLPDDVMGDMKIRSLAREGVIGSFA--WVDPG 97
DB 71 DEIILPDDDEVFLHPCGELALVLTLSVTLPDDLGVGLDGRSSLARLGLMVHTAHRIDPG 130
QY 98 WDNNTLMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYR-----GNVQGS 145
DB 131 WHGQIVLEFYNGKPLALRPGKVGICALSFEPLSGSADRPYNNRQDAKYKNOQGA 185
XX
RESULT 14
ADL05995
ID ADL05995 standard; protein; 192 AA.
XX
XX ADL05995;
XX
XX 06-MAY-2004 (first entry)
XX
XX M. catarrhalis protein #1761.
XX
XX Moraxella catarrhalis; infection.
XX
XX Moraxella catarrhalis.
XX
XX US6673910-B1.
XX
XX 06-JAN-2004.
XX
XX
XX 04-APR-2000; 2000US-00540236.
XX
XX 08-APR-1999; 99US-0128416P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL;
XX
XX WPI; 2004-178127/17.
XX
XX N-PSDB; ADL04075.
XX
XX New nucleic acid encoding a Moraxella catarrhalis polypeptide, useful for
XX preparing a composition for diagnosing, preventing or treating infection
XX caused by Moraxella catarrhalis.
XX
XX Disclosure; SEQ ID NO 3681; 429pp; English.
XX
XX The invention relates to an isolated nucleic acid encoding an Moraxella
XX catarrhalis polypeptide. The nucleic acid is useful for preparing a
XX composition for diagnosing, preventing or treating infection caused by
XX Moraxella catarrhalis. The present sequence represents the amino acid
XX sequence of a M. catarrhalis protein.
XX
XX Sequence 192 AA;
XX
XX Query Match 20.9%; Score 168.5; DB 8; Length 192;
XX Best Local Similarity 28.2%; Pred. No. 1.3e-11;
XX Matches 49; Conservative 30; Mismatches 60; Indels 35; Gaps 6;
XX
QY 13 LIEPFSEESLQ-----PAGYDLRVGRE-----AFVKGK-----LID 43
DB 19 MIEPEPQVRYNAQGQKLVSVGTSSYGYDVRCANEKVFVTNVHSAIVDPKNFDDRSFID 78
QY 44 VEKEGVVIPPREVALILTLERIKLPDDVMGDMKIRSLAREGVIGSFAWVDPGWDGNLT 103
DB 79 IIGD-ECIIPNPFALARTVEYFRIPRDVLTICLGKSTYARCGIIVNVTPLEPEWEHVT 137
XX
QY 104 LMLYNASNEPVELRYGERFVQIAFIRLEGP--ARNPYR---GNVQGSRLAFSK 152
DB 138 LEFSNTTNLPARIYAGEGVAQMLFFQSDADDVCCISYKDRGKYQGQGTGVTLPK 191
XX
RESULT 15
ADFO4778
ID ADFO4778 standard; protein; 204 AA.
XX
XX ADFO4778;
XX
XX 12-FEB-2004 (first entry)
XX
XX Bacterial polypeptide #891.
XX
XX Proteus mirabilis infection; bacterial infection; antibacterial;
XX immunostimulant.
XX
XX Proteus mirabilis.
XX
XX US6605709-B1.
XX
XX 12-AUG-2003.
XX
XX 05-APR-2000; 2000US-00543681.
XX
XX 09-APR-1999; 99US-0128706P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL;
XX
XX WPI; 2003-895291/82.
XX
XX N-PSDB; ADF00606.
XX
```


THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 11:27:13 ; Search time 80.7442 Seconds
(without alignments)
2555.824 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 1998
Sequence: 1 MLHHVKLIYATSKRLVGKK.....MKKKRELARIWDEIBKXLS 403

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1833.5	91.8	402	Q8X250	Q8X250 pyrococcus
2	1667.5	83.5	401	O59114	O59114 pyrococcus
3	1649.5	82.6	401	Q9V0S3	Q9V0S3 pyrococcus
4	781.5	39.1	466	Q8PUP9	Q8PUP9 methanosarc
5	759.5	38.0	431	Q8TRA0	Q8TRA0 methanosarc
6	735	36.8	404	O28628	O28628 archaeoglob
7	731	36.6	382	Q9HH70	Q9HH70 methanobact
8	731	36.6	386	O27284	O27284 methanobact
9	648.5	32.5	413	Q980S0	Q980S0 sulfolobus
10	636.5	31.9	412	O976C4	O976C4 sulfolobus
11	627.5	31.4	403	1 DFP METJA	O58323 m coenzyme
12	626	31.3	392	Q9HRS1	Q9HRS1 halobacteri
13	605	30.3	398	Q971C8	Q971C8 clostridium
14	577	28.9	395	Q8XJ10	Q8XJ10 clostridium
15	571	28.6	414	Q67PR7	Q67PR7 symbiobacte
16	563.5	28.2	399	Q74E40	Q74E40 geobacter s
17	557	27.9	398	Q8R9S8	Q8R9S8 thermoanaer
18	555	27.8	396	Q6LWU8	Q6LWU8 methanococc
19	543.5	27.2	390	1 CABG BORBU	O51752 b coenzyme
20	542	27.1	401	Q819T8	Q819T8 bacillus ce
21	542	27.1	401	Q6HEU6	Q6HEU6 bacillus th
22	541	27.1	401	Q81WG9	Q81WG9 bacillus an
23	540.5	27.1	397	Q89S04	Q89S04 clostridium
24	540	27.0	401	Q636F7	Q636F7 bacillus ce
25	537	26.9	401	Q732K1	Q732K1 bacillus ce
26	535.5	26.8	388	Q66997	Q66997 aquifex aeo
27	535	26.8	437	Q9YAI0	Q9YAI0 aeropyrum p
28	533	26.7	427	Q8CK27	Q8CK27 streptomyce
29	529	26.5	400	Q8AB12	Q8AB12 bacteroides
30	527	26.4	405	Q827Q1	Q827Q1 streptomyce
31	510	25.5	353	Q652W1	Q652W1 borrelia ga

32 509 25.5 376 2 Q8TXG1
33 507.5 25.4 414 2 Q7MPS9
34 506 25.3 399 2 Q92AI3
35 505 25.3 403 2 Q8ZXM3
36 504.5 25.3 401 2 Q8DDX8
37 502.5 25.2 397 2 Q64S04
38 502.5 25.2 403 2 Q6AMM3
39 500.5 25.1 391 2 Q83D56
40 500.5 25.1 402 2 Q8CXH5
41 498.5 24.9 401 2 Q7P314
42 495.5 24.8 404 2 Q8RTI5
43 495 24.8 399 2 Q71XJ1
44 494.5 24.7 404 2 Q9K3Y4
45 494 24.7 399 2 Q8Y674

Q8txg1 methanopyru
Q7mps9 vibrio vuln
Q92ai3 listeria in
Q8zxm3 pyrobaculum
Q8ddx8 vibrio vuln
Q64s04 bacteroides
Q6amm3 desulfofale
Q83d56 coxiella bu
Q8cxh5 oceanobacil
Q7p314 fusobacteri
Q8rti5 fusobacteri
Q71xj1 listeria mo
Q9k3y4 bacillus ha
Q8y674 listeria mo

ALIGNMENTS

RESULT 1

Q8X250 PRELIMINARY; PRT; 402 AA.
AC Q8X250: Q7LWV5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE DNA flavoprotein Dfp (Dna/pantothenate metabolism flavoprotein).
GN OrderedLocNames=PF1440;
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21664347; PubMed=11782527; DOI=10.1073/pnas.0123727999;
RA Hogrefe H.H., Hansen C.J., Scott B.R., Nielson K.B.;
RT "Archaeal dUTPase enhances PCR amplifications with archaeal DNA
polymerases by preventing dUTP incorporation.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:596-601(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY066006; AAL47573.1; -
DR EMBL; AS010247; AAL81564.1; -
DR HSSP; P30197; 1G63.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC dfp; 1.
SQ SEQUENCE 402 AA; 44721 MW; 9C14A470470CB0D8 CRC64;

Query Match 91.8%; Score 1833.5; DB 2; Length 402;
Best Local Similarity 93.3%; Pred. No. 3.3e-96;
Matches 378; Conservative 6; Mismatches 16; Indels 5; Gaps 4;
QY 1 MLHHVKLIYATSKRLVGKKI VXXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKII 59
DB 1 MLHHVKLIYATSKRLVGKKI VLVIAFGSIAAEVCVKLARELIRHGAEVHVMSEAATKII 60
QY 60 HPYAMNLP TGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDDPVPT 119
DB 61 HPYAMFATGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDDPVPT 120
QY 120 TVVTTAFPHIPMIAPAMHETMYRHPVIVRENIERLKKLVGEFTGPIEGRAKVASIDEI 179
DB 121 TVVTTAFPHIPMIAPAMHETMYRHPVIVRENIERLKKLVGEFTGPIEGRAKVASIDEI 180
QY 180 VYRVIKLHKKTLEGRKVLVTAGATREYIDPIRFITNASSGKMGVALAEADFRGA-VTL 238
DB 181 VYRVIKLHKKTLEGRKVLVTAGATREYIDPIRFITNASSGKMGVALAEADFRGA-VTL 240

```
QY 239 IRTKGSVKAIRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKI 298
DB 241 IRTKGSVKSFVENQI--EVEETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKI 298
QY 299 KSGRSITTELVPXNPKIIDRIKEIQPNVFLVGFKAETSKELIEGKQIERAKADLVVG 358
DB 299 KSDRSITTELVP-NPKIIDRIKEIQPNVFLVGFKAETSKELIEGKQIERAKADLVVG 357
QY 359 NTLEAFGSEENQVLIIGRDFTKELPKMKKRELAERIMWDEIEKLS 403
DB 358 NTLEAFGSEENQVLIIGRDFTKELPKMKKRELAERIMWDEIEKLS 402

RESULT 2
O59114 PRELIMINARY; PRT; 401 AA.
AC O59114;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein PH1444.
GN OrderedLocustNames=PH1444;
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo H. Y., Yamazaki J., Kuhlida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AF000006; BAA30551.1; -.
DR PIR; G71018; G71018.
DR HSSP; P30197; IG63.
DR InterPro; IPR005252; Cons hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 401 AA; 44624 MW; 26CB5F523AB02C8 CRC64;

Query Match 83.5%; Score 1667.5; DB 2; Length 401;
Best Local Similarity 83.4%; Pred. No. 8.4e-87;
Matches 337; Conservative 33; Mismatches 29; Indels 5; Gaps 4;

QY 1 MLHVKLIYATSKRLVGKIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEATKII 59
DB 1 MLHVKRIYAKSKRLVGKIVLAIPGSAAEVCKLARELIRHGAHVHVMTPSATKII 60
QY 60 HPYANLPTGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDTPVT 119
DB 61 HPYAMEFATGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDTPVT 120
QY 120 TVVTTAFPHIPIMAPAMHETMYRHPIVRENTERLKKLVGEFIPRIEGRKVASIDEI 179
DB 121 TVVTTAFPHIPIMAPAMHESMYKHPIVRENIEKLKLVGEFIPRIEGRKVASIDEI 180
QY 180 YRVVTKLHKHTLEGKRVLTAGATREYIDPFRITNASSGKGVALAEAEADFRGA-VTL 238
DB 181 YRVVTKLHKHTLEGKRVLTAGATREYIDPFRITNASSGKGVALAEAEADFRGA-VTL 240
QY 239 IRTKGSVKAIRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKI 298
DB 241 IRTKGSVKSFVENQI--EVEETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKI 298
```

```
DB 241 IRTKGSVNSFVENQI--QVETVEEMLAAIEKELTEKDYDVVIMAAVSDFKPKIKAEDKI 298
QY 299 KSGRSITTELVPXNPKIIDRIKEIQPNVFLVGFKAETSKELIEGKQIERAKADLVVG 358
DB 299 KSNKSITTELVP-NPKIIDRIKEIQPNVFLVGFKAETMEKLIQAKKQIERAKSDIVIG 357
QY 359 NTLEAFGSEENQVLIIGRDFTKELPKMKKRELAERIMWDEIEKXL 402
DB 358 NTLEAFGSDSKVLIIGNFEKELPKMKKRELAERIMWDEIEKRL 401

RESULT 3
Q9V0S3 PRELIMINARY; PRT; 401 AA.
AC Q9V0S3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Dfp DNA/pantothenate metabolism flavoprotein.
GN ORFNames=PAB1897;
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RX MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL MBL; Microbiol. 47:1495-1512(2003).
DR EMBL; AJ248285; CAB49630.1; -.
DR PIR; E75114; E75114.
DR HSSP; P30197; IG63.
DR InterPro; IPR005252; Cons hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 401 AA; 44290 MW; 96CCD5A190C1FA5D CRC64;

Query Match 82.6%; Score 1649.5; DB 2; Length 401;
Best Local Similarity 82.9%; Pred. No. 8.8e-86;
Matches 335; Conservative 31; Mismatches 33; Indels 5; Gaps 4;

QY 1 MLHVKLIYATSKRLVGKIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEATKII 59
DB 1 MLSHKLIVATSKRLVGKIVLAIPGSAAEVCKLARELIRHGAHVHVMSPATKII 60
QY 60 HPYANLPTGNPVITEITGFIHVELAGEHENKADLILVCPATANTISKIACGIDTPVT 119
DB 61 HPYAMEFATGNPVITEITGSIHVELAGEHENKADLILVCPATANTISKIACGIDTPVT 120
QY 120 TVVTTAFPHIPIMAPAMHETMYRHPIVRENTERLKKLVGEFIPRIEGRKVASIDEI 179
DB 121 TVVTTAFPHIPIMAPAMHESMYKHPIVRENIEKLKLVGEFIPRIEGRKVASIDEI 180
QY 180 YRVVTKLHKHTLEGKRVLTAGATREYIDPFRITNASSGKGVALAEAEADFRGA-VTL 238
DB 181 YRVVTKLHKHTLEGKRVLTAGATREYIDPFRITNASSGKGVALAEAEADFRGA-VTL 240
QY 239 IRTKGSVKAIRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEGKI 298
DB 241 IRTKGSVKSFVENQI--EVEETVEEMLGAIEBELTKKYDVVIMAAVSDFKPKIKAKKI 298
QY 299 KSGRSITTELVPXNPKIIDRIKEIQPNVFLVGFKAETSKELIEGKQIERAKADLVVG 358
DB 299 KSDKSITTELVP-NPKIIRKIKIEIQSDVFLVGFKAETSMKELISEAKKQIESAGSDLVIG 357
```


QY 359 NTLEAFSGSENOVVLIGRDTFKLPKMKRELAERIWDIEIKXL 402
DB 358 NTLEAFSGSEKVVIVGKOPVKLPKMKRELAERIWDIEIKKI 401

RESULT 4
Q8PUP9

ID Q8PUP9 PRELIMINARY; PRT; 466 AA.
AC Q8PUP9; 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE DNA/pantothenate metabolism flavoprotein.
GN OrderedLocusNames=NM2283;
OS Methanosarcina mazel (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobria; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goel / GoI / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wierzer A., Baumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazel: evidence for lateral gene
RT transfer between Bacteria and Archaea";
RL J. Mol. Microbiol. Biotechnol. 4:453-461 (2002).
RL EMBL; AS013470; AM31979.1; -;
DR HSSP; Q9RC23; 1P3Y.
DR InterPro; IPR005252; Cons hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 466 AA; 50459 MW; 7290D4F35FC5AA25 CRC64;

Query Match 39.1%; Score 781.5; DB 2; Length 466;
Best Local Similarity 45.0%; Pred. No. 1.7e-36;
Matches 186; Conservative 70; Mismatches 136; Indels 21; Gaps 13;
QY 2 LHHVKLIYATKSKRLVGKIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKIIH 60
DB 59 IHPTLWIOGSSSLAGKTVILVGTSGIGAVRVVLSRELIRNGAEVHVMSEAATKIIH 118
QY 61 PYAWNLPNGPVITEITGFIEHVELAGEHENKADLILVCPATANTISKIACGIDDTPTVT 120
DB 119 PDALHYATGNPVAELGGRVHEVFCG-FKGRADLLLIAPATANTIGKIAYGIDDTTTS 177
QY 121 VTTAP-PIPIPIAMHETMYRHPVIRENIRLKLGVFEFGPIREGRKAVASIDEI 179
DB 178 FATTALGSDVPVIVPAMHESMYRHPAVTENIARLKSWMGSMVGPREEGVAKIAGNEE 237
QY 180 VYRVIKLHKHTLEGRKVLVTAGATREYIDPIREITNASSGKMGVALAEADPRGA-VTL 238
DB 238 VLEVERALGDRSLNKVITGGTASLSDPIRLVTRNASKGKTGRELALEYRGADVT 297
QY 239 I-RTKGSVKAFIRIKLKVETVEEMLSAIENELRSKKYDVVIMAAVSDPRPKIKAEG 297
DB 298 VHRDRIGFAGIR---EIFAESAAMEAVLSEL-EKGYDVLISAAIADYTAEPSP-K 351
QY 298 IKSGRSITIELVPXNPKIIDRIKEIOPNVFLVGFKAET--SKKLIEEGKQIERAKADL 355
DB 352 IKSGGFILKLPKTR-KLIKECREKYPDLVIGFKAETGIEKELLRRAAATLEGTLD 410
QY 356 VVGNTL--EAFSGSENOVVLIGRDTFKLPKMK--KRELAERIWDIEIKXL 403
DB 411 IANDVKGKMGMTTEENELLLGK--KSGPRHVSNGNRKRLAASILEVSEILN 461

RESULT 5
Q8TRA0

ID Q8TRA0 PRELIMINARY; PRT; 431 AA.
AC Q8TRA0; 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Pantothenate metabolism flavoprotein.
GN Names=dpf; OrderedLocusNames=MA1280;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobria; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11922238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity";
RL Genome Res. 12:532-542 (2002).
DR EMBL; AE010797; AAM04699.1; -;
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 431 AA; 46193 MW; 646596DDC9E4C156 CRC64;
Query Match 38.0%; Score 759.5; DB 2; Length 431;
Best Local Similarity 43.5%; Pred. No. 2.8e-35;
Matches 178; Conservative 78; Mismatches 136; Indels 17; Gaps 12;
QY 3 HHVKLIYATKSKRLVGKIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKIIHP 61
DB 27 HPTLWIOGKSSLSKTVILVGTSGIGAVRVVLSRELIRNGAEVHVMSEAATKIIHP 86
QY 62 YAWNLPNGPVITEITGFIEHVELAGEHENKADLILVCPATANTISKIACGIDDTPTVT 121
DB 87 DALQYATGNPVTIELGGRVHEVFCG-FKGRADLLLIAPATANTIGKIACGIDDTTTSF 145
QY 122 VTTAP-PIPIPIAMHETMYRHPVIRENIRLKLGVFEFGPIREGRKAVASIDEI 180
DB 146 ATTALGSGVPLVMPVAMPHESMYRHPAVTENIARLKSWMGSMVGPREEGVAKIAGNEE 205
QY 181 YRVIKLHKHTLEGRKVLVTAGATREYIDPIREITNASSGKMGVALAEADPRGA-VTL 239
DB 206 LEVERALGDRSLNKVITGGTASLSDPIRLVTRNASKGKTGRELALEYRGADVT 265
QY 240 -RTKGSVKAFIRIKLKVETVEEMLSAIENELRSKKYDVVIMAAVSDPRPKIKAEG 298
DB 266 HRDRILGKLPKTR-KLIKECREKYPDLVIGFKAETGIEKELLRRAAATLEGTLD 319
QY 299 KSGRSITIELVPXNPKIIDRIKEIOPNVFLVGFKAET--SKKLIEEGKQIERAKADL 356
DB 320 KSGGEFVLKLPKTR-KLIKECREKYPDLVIGFKAETGIEKELLRRAAATLEGTLD 378
QY 357 VVGNTL--EAFSGSENOVVLIGRDTFKLPKMK--KRELAERIWDIEIKXL 403
DB 379 AANDVKGKMGMTTEENELLLGK--KSGPRHVSNGNRKRLAASILEVSEILN 426

```
RESULT 6
O28628      PRELIMINARY;          PRT;    404 AA.
AC O28628;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Pantothenate metabolism flavoprotein (Dfp).
GN OrderedLocusNames=AF1645;
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475; DOI=10.1038/37052;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M.L., Hickey E.K., Peterson J.D.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.R.,
RA Kirnesen E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.J.,
RA Peterson S.N., Reich C.I., McNeil L.K., Badger J.H., Glodek A.,
RA Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L.A.,
RA Uitterback T.R., Cotton M.D., Spriggs T., Artiach P., Kaine B.P.,
RA Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C.,
RA Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O.,
RA Woese C.R., Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370 (1997).
DR EMBL; AE000989; AAB89597.1; -.
DR PIR; D69455; D69455.
DR TIGR; AF1645; -.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
KW Complete proteome.
SQ SEQUENCE 404 AA; 44885 MW; 5A2B899C3E6A66F CRC64;

Query Match      36.8%; Score 735; DB 2; Length 404;
Best Local Similarity 45.0%; Pred. No. 6.4e-34;
Matches 182; Conservative 74; Mismatches 122; Indels 26; Gaps 13;
QY 4 HVKLIYATKSRKLVGKKIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAAATKIHPY 62
DB 5 HLERIRGRSRKLERKKIVLGVGTGSIAAVETVKLARELVRRGADVAVMSRAARKIHPY 64
QY 63 ANWLPTGNPVITEIGTPIHVELAGHENKADLLVCPATANTISKIAGCIDTPTVTTV 122
DB 65 ALEFATGKRVVTEITGSIEHVNLLGEY-GDADFLIAPCTANTISKIAGCIDTPTTFA 123
QY 123 TTAAP-PHIPIMTAPAMHETMYRHPVIRENIRLKLGVFIPRIEGRKAVASIDEIVY 181
DB 124 TTLGSGKPIIIVPAHMEAMKAVLENIQRLIDMGIEFVQPIREEGKAKFPSTETCL 183
QY 182 RVIKHLHKLTKRGKRVLTAGATREYIDPPIRFTNASSGKMGVALAEADPRGA-VTLIR 240
DB 184 HVERELYPKEMGKRVVTSVSGTTEQIDPIRISNKGSRMGLEALEFWRGADVHT 243
QY 241 TKGSKVAFIRIKLVKQVTEVEMLSAENELRSKKYDVVMAAASVDFRPKIKAGKIKS 300
DB 244 SKPSGMSLPNPK-EIRVMSVDMKAVLYEI-GKGCDFVSSAAAADFTVDAEAK-KIKT 300
QY 301 GRSITIELVPXPKIIDRIKEIQPNVFLVGFKAET--SKEKLIIEGKQIERAKADLVVG 358
DB 301 APELVKL-KESPKIIEVKRKIYSG-HIIGFKAETGMSDELKVKASEKKADNLMWVA 358
QY 359 NTL--EAFGSBENQVLIQRDFTKLPKMK-----KRELABRI 394
```

```
Db 359 NDVLERGMGTEDTRVLIL-----TPKRQEWVEGLKQHVABRI 395

RESULT 7
Q9HH70      PRELIMINARY;          PRT;    382 AA.
AC Q9HH70;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE MTW1216.
GN Name=mtw1216;
OS Methanobacterium wolfei.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanobacter.
OX NCBI_TaxID=145261;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=21429258; PubMed=11544247;
RX DOI=10.1128/JB.183.19.5788-5792.2001;
RA Luo Y., Pfister P., Leisinger T., Wasserfallen A.;
RT "The Genome of Archaeal Prophage pafM100 Encodes the Lytic Enzyme
RT Responsible for Autolysis of Methanothermobacter wolfeii."
RL J. Bacteriol. 183:5788-5792 (2001).
RN [2]_
RP SEQUENCE FROM N.A.
RA Luo Y.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF301375; AAG39941.1; -.
DR HSSP; Q9HAB8; IP90.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coaBC_dfp; 1.
SQ SEQUENCE 382 AA; 41207 MW; 8E792042DFDFA3AB CRC64;

Query Match      36.6%; Score 731; DB 2; Length 382;
Best Local Similarity 45.5%; Pred. No. 1e-33;
Matches 178; Conservative 65; Mismatches 126; Indels 22; Gaps 13;
QY 20 KIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAAATKIHPYAMNLPVTEITG 78
DB 2 RVLVCVTGSVAAVEAVKLARELRKGAEBKCFMSDADRIIHPYAMEFATGSPVLELTG 61
QY 79 PIEHVELAGHENKADLLVCPATANTISKIAGCIDTPTVTTVTTPPH-IPIMAPAM 137
DB 62 BIEHVKY-----SDADLLVAPATANIIGKLAYKLADNPISLLLTAAAGRTAILMVP 116
QY 138 HETMYRHPVIRENIRLKLGVFIPRIEGRKAVASIDEIVYRVIKHLHKLTKLEGRV 197
DB 117 HEAMYE--AAENIRKLKEGVAFLEPRMDEGKAKFPDITITLEALROASGKMRGLV 174
QY 198 LVTAGATREYIDPPIRFTNASSGKMGVALAEADPRGA-VTLIRTKGSVK-AFRIRKIKL 255
DB 175 LVSLGGTLEPIDPVRVITNRSSGRMGLAVAREAYIQGADVTLVAGTVSDIPSQRTV-- 232
QY 256 KQVTEVEMLSAENELRSKKYDVVMAAASVDFRPKIKAGKIKSGRSITIELVPXPKI 315
DB 233 RAETAHAEAAVAELI--GEHDVFVSAASVDFRP-VYSEEEKISSDSEITLRLKP-NPKI 288
QY 316 IDRIKEIQPNVFLVGFKAET--TSKEKLIIEGKQIERAKADLVVGN--TLEAFGSBENQV 371
DB 289 IRMARETHPEAFIVGFKAHGVSEBELAAARKQIEDSVADVADVANDVSVEGSGSENNRA 348
QY 372 VLIGRDFTKELPKMKKRELABRIWDEIEKKL 402
DB 349 IIVSGVT-ELPTMKKEELAGLIIGEIMKRL 378

RESULT 8
O27284
```



```
DE Hypothetical protein ST0257.
GN OrderedLocusNames=ST0257;
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Otsuki A., Kosugi H., Hosoyana A., Fukui S.,
RA Nagai Y., Nishijima K., Akeu R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kuehida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagisil M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000982; BAB5223.1; --
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coABC dfp; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 412 AA; 45517 MW; E2B28B9F11338947 CRC64;

Query Match 31.9%; Score 636.5; DB 2; Length 412;
Best Local Similarity 40.6%; Pred. No. 2.5e-28;
Matches 168; Conservative 76; Mismatches 145; Indels 25; Gaps 12;

QY 1 MLHHVLLIYATKSRKLVGKKIVXXPGSTAAAL-DVKACEGLIRHGAHVHVMSEATKII 59
DB 2 VIHPSKKIIGETISSELLNKKILVGVTSVSYKTIIDLVRSLRMLGAEVQVIMSKDASKIV 61

QY 60 HPYANPLPGNVITEITGFIHVELAGHENKADLLVCPATANTISKIAGIDTTPVT 119
DB 62 SPTMFEWATGNPVITKIGGIEHVTLAEDY----DAYVAPATANTISKVNGIADTPT 117

QY 120 TVVTTAFPHI----PTMIAPAMHETVYRHPVIRENIRLKLGVFFIGPRIEGRKAVAS 175
DB 118 ----VTALNFGLKKLLVLPANHLPMYPTPQMINNKLNEIGVIVVEFLVRDVAHPD 174

QY 176 IDEIV-YRVIKLHKHTLGKRLVYTAGATRYIDIPRITWASSGKMGVALAEADFRG 234
DB 175 IEVLSNYISTLLRGKDLGLKIVVTAGPTREHLPVRFISNPSSGTMGVAIANEAYFRG 234

QY 235 A-VTLIRTKGSVKAIRIKLKVETVEMLSAIENELSKKYDVVIMAAVDFRPKIT 293
DB 235 ADVVLVHGLSTHPIKPYVKKVQVETDDEMANEVKKFVET-GYNIVILAGAPADYKFI 293

QY 294 ASGKTKS-GRSTIELVKNPKIIDRIKEIQPNVFLVGFKAETSKS--KLIEGKRQIER 350
DB 294 ASKIDSHTETTVEL-EKTVKISSIRKY--NVFLVGFAATVKKDNLIELKGIKKER 350

QY 351 AKADLVVGNLTLE----AFGSENVQVVLIGRDFTKELPKMKKELAEIRWDETEK 400
DB 351 HGFDLLIANASKDKDIASFSDNEVIVIGNFIKKINKYKTKVIAREILDIVKQ 404

RESULT 11
ID DFP_METJA STANDARD; PRT; 403 AA.
AC Q58323;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Coenzyme A biosynthesis bifunctional protein coABC (DNA/pantothenate
DE metabolism flavoprotein) [includes: Phosphopantothenoylcysteine
DE decarboxylase (EC 4.1.1.36) (PPDC) (CoAC); Phosphopantothenate--
DE cysteine ligase (EC 6.3.2.5) (Phosphopantothenoylcysteine synthase)

DE (PPC synthetase) (PPCS) (CoAB)].
GN Name=coABC; OrderedLocusNames=MJ0913;
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.P., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Catalyzes two steps in the biosynthesis of coenzyme A.
CC In the first step cysteine is conjugated to 4'-phosphopantothenate
CC to form 4'-phosphopantothenoylcysteine, in the latter compound is
CC decarboxylated to form 4'-phosphopantotheine (By similarity).
CC -!- CATALYTIC ACTIVITY: N-[(R)-4'-phosphopantothenoyl]-L-cysteine =
CC pantotheine 4'-phosphate + CO(2).
CC -!- CATALYTIC ACTIVITY: CTP + (R)-4'-phosphopantothenoyl + L-cysteine
CC = CMP + PPI + N-[(R)-4'-phosphopantothenoyl]-L-cysteine.
CC -!- COFACTOR: Binds 1 FMN per subunit (By similarity).
CC -!- PATHWAY: Coenzyme A (CoA) biosynthesis; second step.
CC -!- SUBUNIT: Homodimer; the coAB domains form homodimers.
CC -!- SIMILARITY: In the N-terminal section; belongs to the HPCD (homo-
CC oligomeric flavin containing Cys decarboxylase) superfamily.
CC -!- SIMILARITY: In the C-terminal section; belongs to the PPC
CC synthetase family.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; U67535; AAB98918.1; -.
PIR; A64414; A64414.
HSSP; P30197; 1G63.
TIGR; MJ0913; -.
DR InterPro; IPR011256; Bac_reg_effector.
DR InterPro; IPR005252; Cons_hypoth521.
DR InterPro; IPR007085; Dfp_C.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF04127; DFP; 1.
DR Pfam; PF02441; Flavoprotein; 1.
DR TIGRFAMs; TIGR00521; coABC dfp; 1.
KW Complete proteome; Flavoprotein; FMN; Hypothetical protein; Ligase;
KW Lyase.
FT CHAIN 1 195 Phosphopantothenoylcysteine
FT decarboxylase.
FT CHAIN 196 403 Phosphopantothenate--cysteine ligase.
FT ACT_SITE 87 87 By similarity.
FT SEQUENCE 403 AA; 45670 MW; B8B51AA81A115B94 CRC64;

Query Match 31.4%; Score 627.5; DB 1; Length 403;
Best Local Similarity 39.6%; Pred. No. 8e-28;
Matches 163; Conservative 82; Mismatches 138; Indels 29; Gaps 16;

QY 1 MLHHVLLIYATKSRKLVGKKIVXXPGSTAAALDV-KACEGLIRHGAHVHVMSEATKII 59
DB 5 IMHPTKLLKGTKSLKLENKKILVAVTSSIAAETPKLMRELIRHGAEVYCIITEYTKII 64
```


Search completed: July 2, 2005, 12:24:38
Job time : 83.7442 secs

THIS PAGE TELI REMARK

THIS PAGE LEFT BLANK

Qy 20 KIYXXPGSIALDVKACEGLR---HGAEVHVMSEAATKIHPYAWNLP TGNPVITEI 76

Db 10 KVVGVSGTAA---YKACEILRLITSGCHDVWVPTGSALHFVCAATWSALSGNPVSTEV 67

QY 77 ---TGFIEHVELAGHENKADLLVCPATANTISKIACGIDTPTVTTVTTAFPHI 133
DB 68 WDDVEVPHVR- GQH---ADUVVAPATADMLAAHGLADLLTNTLLTA--RCPVVF 121
QY 134 APAMHETMYRHPVIVRENIERLKKLVGEFIGPR-----IEGRAKV---ASIDEIVYRV 184
DB 122 APAMHETMEHPATQENVATLRRRGALVIEPAVGRLTGVDTKGRLPDPTFEVCRRLV 181
QY 185 KK-LHKHTEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-VTLIRTK 242
DB 182 ARGAAARPDLAGRHHVVSAGGTREPLDPVRFLNRRSSGQGYALARTAAARGARVTLVAAN 241
QY 243 ---GSVKAFIRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAECKI- 298
DB 242 TGLPOLAGVDVVQGTAVQLREAVLKAASDA-----DAVMAAAVADFRPATYATGKIK 295
QY 299 -KSGRSITIELVPXNPKIIDRIK--EIOPNVFLVGFKAETSKBKLIIEGKQIERAKADL 355
DB 296 KDGQEPFAIVLRNPDILAEISADRPGRQVVVGFAAET--DDVLANGRTKLARKGCDL 353
QY 356 VVGNTL---EAFGSEBNOQ-VVLIGRDFTKELPKMKKRELAERIWD 396
DB 354 LVVNEVGERRTGSSENEAVVLGADGSEIPVYPGKEALAEYWD 398

RESULT 2

US-09-738-626-5274
; Sequence 5274, Application US/09738626
; Publication No. US20020197605A1

GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; FILE REFERENCE: 249-125
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5274
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5274

Query Match 22.0%; Score 439; DB 9; Length 420;
Best Local Similarity 31.7%; Pred. No. 4.3e-29;
Matches 132; Conservative 84; Mismatches 147; Indels 54; Gaps 20;

QY 19 KKIVXXPGSIAALDVKACEGLIR----HGAEVHVMSEAATKIIHPYAWNLPNGPVIT 74
DB 11 RNVVGVVAGGIAA--YKACH-IVRAFKEAGDNVRVPTESALKFVGKATFEALSGNPVST 67
QY 75 EI---TGFIHVELAGHENKADLLVCPATANTISKIACGIDD--TPVTTVTTAFPHI 129
DB 69 TVFDAVDSVQHVKGQE-----ADLIVAPATADLMARVVAGRGDDLLAATLLVATC---- 119
QY 130 PIMIAPAMHETMYRHPVIVRENIERLKKLVGEFIGP---RI---EGRKAVASIDEIVYRV 183

DB 120 PVVIAPAMHETMWFNPATVANVATLRQRGITVIEPAHGRRLTGKOTGPGRLDPPEQIV-DL 178
QY 184 IKKHL-----KKTLEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-VT 237
DB 179 ANAVHAGARLQDLAGKKVLITAGTTHEIDPVRFIGNSSSGRQGFALGEIAAQRGAHVS 238
QY 238 LIRTKGSKVAFIRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAECK 297
DB 239 IVAGNAAEELPTPAGAEIVPVVSTQDMFDAVQE--RAGQSDFIWMAAAVADFTPASQATSK 296
QY 298 IKSGRS-----ITIELVPXNPKII-DRIK-----EIOPNVFLVGFKAETSK--KLIEE 343
DB 297 LKKGSDSDSDALSTISLV-ENPDILATTVKREAGELDSNPVIVGFAAETGDEHTTALLEY 355
QY 344 GKROIERAKADLVVGNLT---EAFGSEBNOQ-VVLIGRDFTKELPKMKKRELAERIWD 396
DB 356 ARKLLQKKGCDLLMCNEVGKVKFGQKINEGHWILDAGGVVDVVEHGSKIEVAQAQIWD 412

RESULT 3

US-10-494-675-98
; Sequence 98, Application US/10494675
; Publication No. US20050019877A1

GENERAL INFORMATION:

; APPLICANT: Zelder, Oskar
; APPLICANT: Pompejus, Markus
; APPLICANT: Schroder, Hartwig
; APPLICANT: Kroger, Burkhard
; APPLICANT: Kiopprogge, Corinna
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: Genes coding for metabolic pathway proteins
; FILE REFERENCE: BGI-163US
; CURRENT APPLICATION NUMBER: US/10/494,675
; CURRENT FILING DATE: 2004-05-04
; PRIOR APPLICATION NUMBER: PCT/EP02/12141
; PRIOR FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: DE 101 54 292.1
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 164
; SEQ ID NO 98
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-494-675-98

Query Match 22.0%; Score 439; DB 17; Length 422;
Best Local Similarity 31.7%; Pred. No. 4.3e-29;
Matches 132; Conservative 84; Mismatches 147; Indels 54; Gaps 20;

QY 19 KKIVXXPGSIAALDVKACEGLIR----HGAEVHVMSEAATKIIHPYAWNLPNGPVIT 74
DB 13 RNVVGVVAGGIAA--YKACH-IVRAFKEAGDNVRVPTESALKFVGKATFEALSGNPVST 69
QY 75 EI---TGFIHVELAGHENKADLLVCPATANTISKIACGIDD--TPVTTVTTAFPHI 129
DB 70 TVFDAVDSVQHVKGQE-----ADLIVAPATADLMARVVAGRGDDLLAATLLVATC---- 121
QY 130 PIMIAPAMHETMYRHPVIVRENIERLKKLVGEFIGP---RI---EGRKAVASIDEIVYRV 183
DB 122 PVVIAPAMHETMWFNPATVANVATLRQRGITVIEPAHGRRLTGKOTGPGRLDPPEQIV-DL 180
QY 184 IKKHL-----KKTLEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-VT 237
DB 181 ANAVHAGARLQDLAGKKVLITAGTTHEIDPVRFIGNSSSGRQGFALGEIAAQRGAHVS 240
QY 238 LIRTKGSKVAFIRIKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAECK 297
DB 241 IVAGNAAEELPTPAGAEIVPVVSTQDMFDAVQE--RAGQSDFIWMAAAVADFTPASQATSK 298
QY 298 IKSGRS-----ITIELVPXNPKII-DRIK-----EIOPNVFLVGFKAETSK--KLIEE 343
DB 299 LKKGSDSDSDALSTISLV-ENPDILATTVKREAGELDSNPVIVGFAAETGDEHTTALLEY 357

Db 67 IREITNTKOLL--IEMQERVQDYQVLIHSMVSDYTPVYMTGLVEEQVQSSNLKEFLSKQN 124

Qy 294 AEGKIKSGRSITELVPXNPKIIDRIKBIQPNVFLVGFK--AETSKEKLIBEGKQIERA 351

Db 125 HQAKISSTDEBVQLFKTKPKIISLVKEMNPTIHLIGFKLLVDVTDHLVDIARKSLIKN 184

Qy 352 KADLVVGNLTLEAFSGSEENVVLIGROFTKELPKMKKRELAERIWDEIE 399

Db 185 QADLIITANDLTQISADQHRAIFVEKNQQTQV--QTKEEIAELLELEKIQ 230

RESULT 9

US-10-617-320-4272

; Sequence 4272, Application US/10617320

; Publication No. US20050136404A1

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID

; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC THERAPEUTICS

; NUMBER OF SEQUENCES: 5206

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: <Unknown>

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: <Unknown>

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/617,320

; FILING DATE: 10-Jul-2003

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,433

; FILING DATE: 30-Jun-1998

; APPLICATION NUMBER: 60/085131

; FILING DATE: May 12, 1998

; APPLICATION NUMBER: 60/051553

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-011

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 4272:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 154 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ORIGINAL SOURCE:

; ORGANISM: Streptococcus pneumoniae

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (B) LOCATION 1...154

; SEQUENCE DESCRIPTION: SEQ ID NO: 4272:

US-10-617-320-4272

Query Match 11.0%; Score 219; DB 18; Length 154;

Best Local Similarity 36.9%; Pred. No. 9.6e-11;

Matches 52; Conservative 25; Mismatches 54; Indels 10; Gaps 3;

Qy 21 IVXXPGSIAAL-DVKACEGLIRHGAHVHVMSEAATKIHPYAWNLPTGNPVITEI--- 76

Db 16 ILLAVTGSIASYKSAADLVSSLKKQGHQVTVLMTQAATFEIQPLTLQVLSQNPVHLDVMKE 75

Qy 77 --TGFIEHVELAGEHENKADLILVCPATANTISKIACIGDDTPTVTVTTFAPFHIPIMIA 134

Db 76 PYPDQVNHIELG---KKADLFLIVWPATANTIAKLAHGFADNMVSTALALPSPHIA 131

Qy 135 PAMHETMYRHPVIRENIERLK 155

Db 132 PAMNTQYDHPVTQNNLKTLE 152

RESULT 10

US-09-895-913A-66

; Sequence 66, Application US/09895913A

; Patent No. US20020160456A1

; GENERAL INFORMATION:

; APPLICANT: Kleanthous, Harold

; APPLICANT: Al-Garawi, Amal

; APPLICANT: Miller, Charles

; APPLICANT: Tomb, Jean Francois

; APPLICANT: Oomen, Raymond P.

; TITLE OF INVENTION: Identification of Polynucleotides

; TITLE OF INVENTION: Encoding No. US20020160456A1el Helicobacter Polypeptides in the

; FILE REFERENCE: 06132/043002

; CURRENT APPLICATION NUMBER: US/09/895,913A

; CURRENT FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 08/881,227

; PRIOR FILING DATE: 1997-06-24

; NUMBER OF SEQ ID NOS: 368

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 66

; LENGTH: 228

; TYPE: PRT

; ORGANISM: Helicobacter pylori

US-09-895-913A-66

Query Match 9.9%; Score 197; DB 9; Length 228;

Best Local Similarity 28.3%; Pred. No. 1.3e-08;

Matches 63; Conservative 44; Mismatches 88; Indels 28; Gaps 7;

Qy 14 RKLVGKKIVXXPGSIAAL-DVKACEGLIRHGAHVHVMSEAATKIHPYAWNLPTGNPV 72

Db 12 RULENKRVLVLSGSGIAAYKSLLEVLFLFKSGASIQVVMKGAKKFKPLSPALSHHKV 71

Qy 73 ITEITGFIHVELAGEHEN-----KADLILVCPATANTISKIACIGDDTPTVT-T 120

Db 72 LHD-----RNEKYYVNHQNALHHNHIAACAANADLLIFAPLSTNSLSKIAHALADNIVSAT 126

Qy 121 VTTAPPHIPIMTAPAMHETMYRHPVIRENIERLKLGVEFFIGPR-----IEEGRAKVA 174

Db 127 FLACASPKI---LAPSMNTNMLNSPTQSNLKRKDSNHIILDTKNALLACDTKGGGAMA 183

Qy 175 SIDEIVYRVIKLHKKT-LEGKRVLVTAGATREYIDPIRFITN 216

Db 184 EPLEILFKAAQTLLKDAYFENREIVMGASIEKIDSVRTISN 226

RESULT 11

US-10-276-774-2001

; Sequence 2001, Application US/10276774

; Publication No. US20040053245A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; APPLICANT: Tang, Y, Tom et al

; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides

; FILE REFERENCE: 21272-030

; CURRENT APPLICATION NUMBER: US/10/276,774

; CURRENT FILING DATE: 2002-11-18

; PRIOR APPLICATION NUMBER: 09/560,875

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: 09/496,914

; PRIOR FILING DATE: 2000-02-03

; NUMBER OF SEQ ID NOS: 2700

; SOFTWARE: Custom

Qy	9	YATKSRKLVGKKIVXXPGS	AAALDVKACBGLIRHGA	EVHVMSEAA	TKIIH-PYANWLP	67
		:	:	:	:	
		:	:	:	:	
Db	15	YQPSR----	PRVLLAASG	VAAIKF----	ESLCRSFSEWADVRAV	ATSSLHFVDRSSLP 67
		:	:	:	:	
Qy	68	TGNPVITE-----	ITGFIEHVELAGEHEN	KADLILVCP	DATATISKIACID	DDTPV 118
		:	:	:	:	
Db	68	SGIVLYTDDDEW	STWKKGID	VELHIEL----	RKWADVWVIAP	SANTLAKIAGSLCDNLL 123
		:	:	:	:	
Qy	119	TTVVTAPPHIPIM	TAPAMHETWVRHP	IVRENIER	KLKGV	FEFTGP 164
		:	:	:	:	
Db	124	TCIVRAWDYSK	FLFVAPAMNTLMN	NPPTFESHLHT	INOLG	IALIPP 165
		:	:	:	:	
		:	:	:	:	

RESULT 15

```

US-10-425-115-322885
; Sequence 322885, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 322885
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_57534C.1.pep
US-10-425-115-322885

```

Query Match . 8.1%; Score 161.5; DB 16; Length 220;
Best Local Similarity 28.3%; Pred. No. 1.4e-05;
Matches 47; Conservative 35; Mismatches 63; Indels 21; Gaps 5;

		9	YATGRKVLGGKIVXXVPGSTAAALDVKACEGLIRHGAEVHAVMSEAA TKIIH--PYAWNLP	67
Qy	:		: :	
		15	YQPGR-----PRVLLAASGSAAIKF---ESLCRSFSEWADVRAVATTSSLHFVDRSSLP	67
Db	:		: :	
		68	TGNPVITE-----ITGFIEHVLAGEHENKADILVCPATANTISKIAGIDTTPV	118
Qy	:		: :	
		68	SGIVLTDDDEWSWKKGIDVELHIEL-----RKWADVVIAPLSANTLAKIAGSLCDNLL	123
Db	:		: :	
		119	TTVTVPAPHPIMTAPAMHETWVRHPVIRENIERKLKLVGEFTGP	164
Qy	:		: :	
		124	TCIGVADWSYKSLFFVAPAMNTLMWNPPTESHLHTINOLGIATIPP	165
Db	:		: :	

Search completed: July 2, 2005, 12:34:04
Job time : 81.0233 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 11:06:33 ; Search time 85.0698 Seconds
(without alignments)
1832.195 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 1998
Sequence: 1 MLHHVKLIYATSKRLVGKK.....KKMKRELAEIRINDEIEKXLS 403

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1980s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1985	99.3	403	2 AAW72844	AAW72844 Polymeras
2	1649.5	82.6	401	4 AAB96653	AAB96653 Putative
3	1574.5	78.8	406	8 ADN46559	ADN46559 Thermococ
4	509	25.5	376	7 ADM26107	ADM26107 Hyperther
5	494	24.7	399	5 ABB47692	ABB47692 Listeria
6	489	24.5	405	7 ABO75311	ABO75311 Pseudomon
7	466	23.3	418	7 ADF07150	ADF07150 Bacterial
8	461.5	23.1	399	4 AAG82741	AAG82741 S. epider
9	461.5	23.1	400	5 ABP39940	ABP39940 Staphyloc
10	461.5	23.1	400	8 ADS06229	ADS06229 Staphyloc
11	458.5	22.9	431	8 ADL05850	ADL05850 M. catarr
12	457	22.9	394	6 ABP80942	ABP80942 N. gonorr
13	453.5	22.7	401	6 ABM71249	ABM71249 Staphyloc
14	444	22.2	401	6 ADB10278	ADB10278 Alloiococ
15	444	22.2	401	8 ADJ27133	ADJ27133 Alloiococ
16	444	22.2	422	4 AAB79946	AAB79946 Corynebac
17	439	22.0	420	4 AAG91520	AAG91520 C. glutami
18	439	22.0	422	7 ADD13531	ADD13531 C. glutami
19	439	22.0	427	6 ADA33918	ADA33918 Acinetoba
20	424.5	21.2	430	6 ABR63491	ABR63491 E. coli fl
21	424.5	21.2	430	6 ABR63492	ABR63492 E. coli fl
22	422	21.1	414	6 ABM65535	ABM65535 Propionib
23	418.5	20.9	398	7 ABO64139	ABO64139 Klebsiell
24	410	20.5	428	6 ABM68817	ABM68817 Phototrab
25	264	13.2	184	8 ADS51776	ADS51776 Pantothene

26	262	13.1	183	8 ADK48063	ADK48063 Streptoco
27	260	13.0	183	6 ABU01661	ABU01661 S. pneumo
28	260	13.0	183	6 ABP81597	ABP81597 Streptoco
29	260	13.0	184	8 ADS51774	ADS51774 Pantothene
30	254.5	12.7	299	5 ABU51464	ABU51464 Helicobac
31	244	12.2	180	5 ABP28493	ABP28493 Streptoco
32	244	12.2	230	5 ABP30330	ABP30330 Streptoco
33	244	12.2	233	5 ABP28495	ABP28495 Streptoco
34	240	12.0	304	4 AAU67705	AAU67705 Propionib
35	240	12.0	304	6 ABM64224	ABM64224 Propionib
36	238	11.9	234	5 ABU01660	ABU01660 S. pneumo
37	237	11.9	230	5 ABP28496	ABP28496 Streptoco
38	236	11.8	252	8 ADK48064	ADK48064 Streptoco
39	235.5	11.8	245	7 ADC94821	ADC94821 E. faeciu
40	235	11.8	188	7 ADC97191	ADC97191 E. faeciu
41	233	11.7	262	7 ADH86020	ADH86020 Enterococ
42	232	11.6	181	5 ABP28494	ABP28494 Streptoco
43	224.5	11.2	231	5 ABB53883	ABB53883 Lactococc
44	223	11.2	178	5 ABB53882	ABB53882 Lactococc
45	219	11.0	154	8 ADR95637	ADR95637 Novel S.

ALIGNMENTS

RESULT 1
AAW72844
ID AAW72844 standard; protein; 403 AA.
XX
AC AAW72844;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P50 component.
XX
KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
KW sequencing; replication.
XX
OS Pyrococcus furiosus; strain DSM 3638.
XX
FH Key Location/Qualifiers
FT Peptide 1..31
FT Misc-difference 23 /note= "N-terminal peptide used to generate primers"
FT Misc-difference 24 /note= "encoded by NNN"
FT Misc-difference 25 /note= "encoded by NNN"
FT Peptide 275..291 /note= "internal peptide used to generate primers"
FT Misc-difference 310 /note= "encoded by CCN"
FT Misc-difference 311 /note= "encoded by NNN"
FT Peptide 353..376 /note= "internal peptide used to generate primers"
XX
WO9842860-Al.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US005497.
XX
PR 21-MAR-1997; 97US-00822774.
PR 24-OCT-1997; 97US-00957709.
XX
(STRA-) STRATAGENE.
XX
PI Hogrefe H, Hansen CJ;
XX
DR WPI; 1998-542284/46.

DR N-PSDB; AAV63859.
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 XX
 PS Claim 46; Page 36; 161pp; English.
 XX
 CC This is the amino acid sequence of the P50 component of the polymerase
 CC enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. The sequence is
 CC predicted from a DNA sequence (see AAV63859) obtained from genomic DNA by
 CC PCR. P50 and P45 (see AAW72847) are the predominant components of PEF,
 CC which acts to enhance the activity of P. furiosus DNA polymerase, thereby
 CC providing replication products of greater length and purity. P50 is
 CC similar in structure to a bacterial flavoprotein. The invention provides
 CC novel extracts, proteins and complexes that improve the polymerisation
 CC activity of nucleic acid polymerases, as well as DNA constructs and
 CC antibodies raised against P45 or P50. Also included are methods for
 CC identifying compositions with polymerase enhancing activity, for
 CC purifying and using these compositions, and specific extracts, proteins
 CC and complexes that function to enhance polymerase activity. Nucleic acid
 CC polymerase reactions can be enhanced (claimed) by mixing a nucleic acid
 CC template, at least 1 polymerase and a composition having polymerase
 CC enhancing activity. Kits are provided for replicating nucleic acids. The
 CC kits can be used in site-directed mutagenesis, nucleic acid sequencing or
 CC amplification (preferably PCR or RT-PCR). (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX
 XX Sequence 403 AA;

Query Match 99.3%; Score 1985; DB 2; Length 403;
 Best Local Similarity 99.8%; Pred. No. 4.3e-184; Mismatches 0; Gaps 0;
 Matches 402; Conservative 1; Indels 0;
 QY 1 MLHHVKLIYATSKRLVGKKIVXXPGSIAALDVKACEGLIRHGAHVHVMSEAAATKIIH 60
 DB 1 MLHHVKLIYATSKRLVGKKIVXXPGSIAALDVKACEGLIRHGAHVHVMSEAAATKIIH 60
 QY 61 PYAWNLPTGNPVITETITGFIHVELAGEHENKADLILVCPATANTISKIACGIDTPVTT 120
 DB 61 PYAWNLPTGNPVITETITGFIHVELAGEHENKADLILVCPATANTISKIACGIDTPVTT 120
 QY 121 VVTTAFPHIPIMAPAMHETMYRHPIVRENIERLKLGVETFGPRIEGRKAVASIDETV 180
 DB 121 VVTTAFPHIPIMAPAMHETMYRHPIVRENIERLKLGVETFGPRIEGRKAVASIDETV 180
 QY 181 YRVIKLHKHTLEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGAVTLIR 240
 DB 181 YRVIKLHKHTLEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGAVTLIR 240
 QY 241 TKGSVKAFIRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEQKIKS 300
 DB 241 TKGSVKAFIRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEQKIKS 300
 QY 301 GRSITIELVPXNPKIITDRIKETOPNVFLVGFKAETSKELIEGKQIERAKADLVVGN 360
 DB 301 GRSITIELVPXNPKIITDRIKETOPNVFLVGFKAETSKELIEGKQIERAKADLVVGN 360
 QY 361 LEAFGSEENQVLIIGRDFTKELPKMKKRELAERIWDEIEKXLS 403
 DB 361 LEAFGSEENQVLIIGRDFTKELPKMKKRELAERIWDEIEKXLS 403

RESULT 2
 AAB96653
 ID AAB96653 standard; protein; 401 AA.
 XX
 XX AAB96653;
 AC
 DT 29-OCT-2001 (first entry)
 XX Putative P. abyssi flavoprotein involved in panthothenate metabolism.
 DE
 XX

KW Hyperthermophilic archaeon; hyperthermophilic protein.

XX Pyrococcus abyssi.

XX FR2792651-Al.

XX 27-OCT-2000.

XX 21-APR-1999; 99FR-00005034.

XX 21-APR-1999; 99FR-00005034.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX (IFRE-) IFREMER INST FR RECH EXPL MER.

XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

XX Querellou J, Weissenbach J, Saurin W, Heilig R;

XX WPI; 2001-126236/14.

XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins
 PT useful in industry.

XX Claim 7; Page 1403-1404; 1657pp; French.

XX The present invention relates to the genomic sequence of *Pyrococcus*
 CC abyssi (see AAF8431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
 CC a hyperthermophilic archaeon, which is isolated from deep-sea
 CC hydrothermal vents. The present invention has one such P. abyssi protein.
 CC The proteins of the present invention have various potential industrial
 CC uses, since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade. Note: This patent is in the same patent family as
 CC WO200005052, which contains additional sequences as shown in AAB99132-
 CC AAB99143, AAH75903-AAH75920 and AAG66436

XX Sequence 401 AA;

Query Match 82.6%; Score 1649.5; DB 4; Length 401;

Best Local Similarity 82.9%; Pred. No. 2e-151;

Matches 335; Conservative 31; Mismatches 33; Indels 5; Gaps 4;

QY 1 MLHHVKLIYATSKRLVGKKIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAAATKII 59
 DB 1 MLSHKLIYATSKRLVGKKIVLAIPGSIAAVECVKLARELIRHGAHVHVMSPSATKII 60
 QY 60 HPYAWNLPTGNPVITETITGFIHVELAGEHENKADLILVCPATANTISKIACGIDTPVT 119
 DB 61 HPYAMEFATGNPVITETITGSIHVELAGEHENKADLILVCPATANTISKIACGIDTPVT 120
 QY 120 TVVTTAFPHIPIMAPAMHETMYRHPIVRENIERLKLGVETFGPRIEGRKAVASIDETV 179
 DB 121 TVVTTAFPHIPIMAPAMHESMYKPIVRENIEKLVGVEFGPRIEGRKAVATIDETV 180
 QY 180 YRVIKLHKHTLEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-VTL 238
 DB 181 YRVIKLHKHTLEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGAETV 240
 QY 239 IRTGSKVAFIRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEQKI 298
 DB 241 IRTGSKVKSFVENQI--EVETVEEMLGAIEBELTKKYDWDVIMAAVSDFRPKIKAEQKI 298
 QY 299 KSGRSITIELVPXNPKIITDRIKETOPNVFLVGFKAETSKELIEGKQIERAKADLVV 358
 DB 299 KSDKSITIELVP-NPKIIRKIKETOSDVLVGFKAETSKELIEGKQIERAKADLVV 357
 QY 359 NTLEAFGSEENQVLIIGRDFTKELPKMKKRELAERIWDEIEKXL 402
 DB 358 NTLEAFGSEESKVVIVGDFVKELPKMKKRELAERIWDEIEKXI 401

RESULT 3
 ADN46559
 ID ADN46559 standard; protein; 406 AA.

XX AC ADN46559;
XX AC
XX DT 01-JUL-2004 (first entry)
XX DE Thermococcus kodakaraensis KOD1 protein sequence SeqID437.
XX DE
XX DE gene disruption; gene targeting; marker gene; transformation;
KW homologous recombination; hyperthermostable archaeobacterium; KOD1;
KW gene structure; gene function; enzyme activity; medicine;
KW forensic science; food; drug inspection; molecular biology; immunology.
XX Thermococcus kodakaraensis.
XX OS
XX PN WO2004022736-A1.
XX PD
XX PD 18-MAR-2004.
XX PF 29-AUG-2003; 2003WO-IB003597.
XX XX
XX PR 30-AUG-2002; 2002JP-00319011.
XX XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX PA Imanaka T, Atomi H;
XX PI
XX XX
XX DR WPI; 2004-257583/24.
XX XX
XX PT Method for disrupting targeted gene in genome of organism particularly
PT thermostable bacterium and with genome chips for analysis, applicable in
PT studying gene structure and functions.
XX XX
XX PS Claim 9; SEQ ID NO 437; 598pp; Japanese.
XX XX
XX CC This invention relates to a novel method for targeting disruption of an
CC arbitrary gene in a genome of an organism which comprises providing the
CC whole sequential data of the genome of such organism, selecting at least
CC 1 arbitrary region in the sequence, providing a vector that contains a
CC sequence homologous with the selected region and a marker gene,
CC transformation, and homologous recombination. The genome is preferably
CC the genome of a hyperthermostable archaeobacterium, particularly
CC Thermococcus kodakaraensis KOD1. The method is for targeting the
CC disruption of a gene in the genome of an organism, which is applicable in
CC studying gene structure and functions as well as enzyme activities of
CC encoded proteins and useful in medicine, forensic science, food or drug
CC inspection, molecular biology and immunology. With this method, the
CC disruption of a gene at an arbitrary position in a genome can be achieved
CC efficiently and reliably. The present sequence is that of a protein
CC encoded by the genome of Thermococcus kodakaraensis which was derived
CC using the method of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX XX
XX SQ Sequence 406 AA;
Query Match 78.8%; Score 1574.5; DB 8; Length 406;
Best Local Similarity 78.8%; Pred. No. 4.1e-144;
Matches 370; Conservative 37; Mismatches 42; Indels 7; Gaps 5;
QY 1 MLHHVKLIYATSRKLVGKKIVXXXPSTAAID-VKACEGLIRHGAHVHVMSEAAATKII 59
DB 4 MLHHVKLIYATSRKLVGKKIVLAIPGSTAAVCEVKLARELIRHGAHVHVMSEAAATKII 63
QY 60 HPYANNLPNGNVITTEITGFIEHVELAGHENKADILVCPATANTISKIAGIDDPVT 119
DB 64 HPYAMEFATGNPVVTEITGFIEHVELAGHENKADILVCPATANTISKIAGIDDPVT 123
QY 120 TVVTTAFPIPIAMPAHETMYRHPIVRENIERLKKLGVEFTGRIEGRKAVASIDRI 179
DB 124 TVVTTAFATPIPIAMPAHETMYRHPIVRENIERLKKLGVEFTGRIEGRKAVASIEI 183
QY 180 VYRVIKKLHKHTLEGKRVLVTAGATREYIDPIRITNASSGKMGVALAEAEADFRGA-VTL 238

DB 184 VYRVIKKLHKHTLEGKRVLVTAGATREYIDPIRITNASSGKMGVALAEAEADFRGA-VTL 243
QY 239 IRTKGSVKAFIRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAAVSDFRPKIKAEKGI 298
DB 244 IRTKSSVPSFVENQI--EVETVEEMLSAIENELRSKKYDVVIMAAAVSDFRPKIKAEKGI 301
QY 299 KSGRITITELVPXPKIIDRIKEIQNPVPLVGFKAET--SKEKLTIEGKQRIERAKADLV 356
DB 302 KSGQPLVLELEP-TPKIIDRVKELOPGVFLVGFKAETGLSEBELTSAARKQIERAGSDLV 360
QY 357 VGNLTLEAFSGSENOVVLIGRDTFKELPKMKKRELAERIWDEIEKXL 402
DB 361 VANTLKAFGSENEVVLVGRDPAKLPRTWKRELAERLWDEIEKXL 406
RESULT 4
ADM26107
ID ADM26107 standard; protein; 376 AA.
XX AC ADM26107;
XX XX
XX DT 20-MAY-2004 (first entry)
XX DE Hyperthermophile Methanopyrus kandleri protein #713.
XX KW hyperthermophile; protein stability enhancement;
KW protein activity enhancement.
XX OS Methanopyrus kandleri.
XX XX
XX PN WO2003076575-A2.
XX PD
XX PD 18-SEP-2003.
XX PF 04-MAR-2003; 2003WO-US006664.
XX PR 04-MAR-2002; 2002US-0361742P.
PR 14-MAY-2002; 2002US-0380423P.
PR 16-SEP-2002; 2002US-0410974P.
XX (FIDE-) FIDELITY SYSTEMS INC.
PA (MALY/) MALYKH A.
XX Slesarev AI, Pavlov A, Pavlova N, Kozyavkin S;
PI WPI; 2003-748383/70.
XX N-PSDB; ADM27081.
DR New isolated nucleic acids encoding any of about 1700 Methanopyrus
PT kandleri proteins, and the encoded proteins, useful as a medicaments or
PT as diagnostic agents.
XX Claim 31; SEQ ID NO 713; 1023pp; English.
XX CC The invention comprises the amino acid sequence of proteins from the
CC hyperthermophile Methanopyrus kandleri, the invention also comprises the
CC complete genome from Methanopyrus kandleri. The Methanopyrus kandleri
CC proteins of the invention are useful for enhancing the stability and/or
CC activity of other proteins. The Methanopyrus kandleri genome is useful in
CC a variety of diagnostic and analytical methods. The present amino acid
CC sequence represents a Methanopyrus kandleri protein of the invention.
XX SQ Sequence 376 AA;
Query Match 25.5%; Score 509; DB 7; Length 376;
Best Local Similarity 37.9%; Pred. No. 2.1e-40;
Matches 149; Conservative 59; Mismatches 149; Indels 36; Gaps 15;
QY 21 IVXXXPSTAAAL--DVKACEGLIRHGAHVHVMSEAAATKIIHPYANNLPNGNPVITEIT 77
DB 4 VVVCVTGSAVVRAPDV--CRELVGRHVRVVAEEALRFVKDGLGF-AAEEVIFRLT 60

```
Qy 78 GFIEHVELAGEHENKADILVCPATANTISKIACGIDDPV--TTVTTAFPHIPIMIAPA 136
Db 61 SRAEHVELA-----EWADVAVVPATLNLAKIARGVADAPVPLTAVTSLGAGKRLVWAPA 116
Qy 137 MHETMYRHPIVRENIERLKKLGEVFIPIGRIEGRKAVASIDEIVYRVIKLHKHKTLEGR 196
Db 117 MSLHMYRPPAREILRQLEDMGVTVVGPVIEGKAKLASIEIEVAV-----EGELSGFR 171
Qy 197 VLVTAGATREYIDPFRFITNASSGKMGVALAEADFRGAVTLIRTKSVKAFRI-----RK 252
Db 172 VLVTGCTMEPLDDVITNRSSGRACEICRELARGA-EVVFVHGFL---RVDPPMD 227
Qy 253 IKLVETVEMLSAENELRSKKYDVWMAAAVSDFRPKIKAEVKIKSGRSITIELVPXN 312
Db 228 ERVEVETREMLEEVVK--RIDVDVAIVMAAAPSDFRPAERADGKLDSSRREHEIRLITE 285
Qy 313 PKIIDRIEIQPNV--FLVGFKAETSKELIBEGKROIERAKADIVGNTLEAFSGEENQ 370
Db 286 -KI---VREVPDFDGVRAFKLDPEP---VEGARLLDEVPCVELVWVNPETAGEGSE 338
Qy 371 -VVLIGRDFTKELPKMKKRELAERIWDIEKXL 402
Db 339 WMLDGDGEVIERVGRDKRELARKLVDALSULL 371
```

RESULT 5

```
AB547692
ID ABB47692 standard; protein; 399 AA.
```

```
XX AC ABB47692;
```

```
XX DT 05-FEB-2002 (first entry)
```

```
XX DE Listeria monocytogenes protein #396.
```

```
XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX KW vitamin B12; bacterial infection; disease.
```

```
XX OS Listeria monocytogenes.
```

```
XX PN WO200177335-A2.
```

```
XX PD 18-OCT-2001.
```

```
XX PF 11-APR-2001; 2001WO-FR001118.
```

```
XX PR 11-APR-2000; 2000FR-00004629.
```

```
XX PA (INSP ) INST PASTEUR.
```

```
XX PI Buchrieser C, Frangeul L, Couve E, Rueniok C, Psihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaeser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-diaz J, Baquero E, Garcia Del Portillo P, Gomez-Lopez N;
PI Madueno E, De Pablo B, Wehlend J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
```

```
XX WPI; 2002-010914/01.
```

```
XX PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX PT and prevention of Listeria and related bacterial infections, and related
XX PT polypeptides.
```

```
XX PS Claim 6; SEQ ID NO 397; 192pp; French.
```

```
XX CC The present invention relates to the genome sequence of Listeria
XX CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
XX CC it are useful for selecting probes and primers for detecting genes in L.
XX CC monocytogenes and related organisms, and for studying genetic
XX CC polymorphisms and other genomes. The present sequence is a protein
XX CC encoded by the genome sequence of the present invention. Proteins
```

```
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccine compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 399 AA;
```

```
Query Match 24.7%; Score 494; DB 5; Length 399;
Best Local Similarity 34.8%; Pred. No. 6.5e-39;
Matches 144; Conservative 70; Mismatches 152; Indels 48; Gaps 16;
```

```
Qy 18 GKIVXXXPGSIAALDVKACEG-LIRHGAEVHVMSEAAKIIHYPANWLPNGNPVITEI 76
Db 3 GRNILLAVSGGIANYKVALTSKLTQAGANVVMWTAHAQEFVPLSFQVLSKNDVYDT 62
Qy 77 -----TGFIEHVELAGEHENKADILVCPATANTISKIACGIDDPVTTVTTAFPHIPI 131
Db 63 FDEKSSVVAHIDLA---DWADLVIVAPATANVIGKMWANGIADDMVTTTILAT--EAPV 116
Qy 132 MIAPAMHETMYRHPIVRENIERLKKLGEVFIPIGRIE-----GRAKVASIDEIVRV 183
Db 117 WVAPAMVHMTQHPAVIRNINRLYADGVRFIEP--EGYLACGYVGRGRLEPEKIVLRI 174
Qy 184 IK--KLHKKTLEGRKVLVTAGATREYIDPFRFITNASSGKMGVALAEADFRGA-VTLIR 240
Db 175 AEFQEDKNLQGNLVLTAGATREKLDPVYFTNHSTGKMGFSAESAARHGANVTLIT 234
Qy 241 TKGS-----VKAFRIKIKLVETVEMLSAENELRSKKYDVVIMAAVSDFRPKKA 294
Db 235 TSKALPVPBGVEA-----IYVESAEEMHQAV-NE-RKVSQDIFVMTAAVADYTPAQVS 285
Qy 295 EGKI-KSGRSITIELVPXNPKIIDRIEIQPNVFLVGFKAETSKELIEEGKROIERAKA 353
Db 286 DQIKKQPGDFTIAMKRTKQILLEGHTSEQVVGFAAET--ENVANARKKUTSKNA 343
Qy 354 DLVVGNTLE----AFGSEENQVLIGRDFTKELPKMKKRELAERIWDIEKXL 402
Db 344 DMIVANNISEAGAGSGDNTIVTFYRKDGSSEALPILDKVEVAEHIKEAANFL 397
```

RESULT 6

```
ABO75311
```

```
XX ID ABO75311 standard; protein; 405 AA.
```

```
XX AC ABO75311;
```

```
XX DT 29-JUL-2004 (first entry)
```

```
XX DE Pseudomonas aeruginosa polypeptide #7486.
```

```
XX KW Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
```

```
XX OS Pseudomonas aeruginosa.
```

```
XX PN US6551795-B1.
```

```
XX PD 22-APR-2003.
```

```
XX PF 18-FEB-1999; 99US-00252991.
```

```
XX PR 18-FEB-1998; 98US-0074788P.
```

```
XX PR 27-JUL-1998; 98US-0094190P.
```

```
XX PA (GENO-) GENOME THERAPEUTICS CORP.
```

PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
DR N-PSDB; ABD0882.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
XX Disclosure; SEQ ID NO 24057; 455pp; English.
XX
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences AB067826-
CC AB084396 represent P. aeruginosa polypeptides of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 405 AA;
Query Match 24.5%; Score 489; DB 7; Length 405;
Best Local Similarity 34.8%; Pred. No. 2e-38;
Matches 136; Conservative 77; Mismatches 132; Indels 46; Gaps 16;
QY 14 RLVGKKIVXXPGSSTAALDVKACEGLIR-----HGAEVHAVMSEATKLIHPYANLPTG 69
DB 5 QRLYRKIRVLGVGGGIAA--YKSAE-LVRLRDQGAERVVMTQGGREITPTLQALSG 61
QY 70 NPVITEI-----TGFIEHVELAGEHENKADLIILVCPATANTISKIACGIDDTPTVTWTT 124
DB 62 HPVHTDLLDPAAEAAMGHIELA---RWADLVLIAPATADLMARLVQGVANDLLTLVLA 117
QY 125 APHPIDMTAPAMHETMYRHPVIRENIERLKKLVGEFIGPRI-----BEGRAKVASIDE 178
DB 118 T--DAQIALAPAMNQAMWRDTATQANAEILLRQRFHLFGPAAGSQACGDVGLGRMLEABE 175
QY 179 IYVRVVKLHKHKKLECKRVLVTAGATREYIDPIRFTLNASSGKMGVALAEADFRGA-VT 237
DB 176 LAQRAADCQORQALTGVHVLITAGTQENIDPVYITNHSNGKMGFALAEAAVEAGARVT 235
QY 238 LIRTKGSV---KAFIRIKIKLVETVEEMLSAIENELRSKKYDVWVMAAAVSDFRPKKA 294
DB 236 LV--TGPVHLPTDPRQVRV--DVVSARDMLAACEAEM---PCDLLASAADYREPVVA 288
QY 295 EGKIK----SGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKELIEGKQRIER 350
DB 289 AHKKKDKPTSGEGLLLQLV-RNPDILATLAQREDRPFVSGFAAET--ENLLDYAARKLKD 345
QY 351 AKADLVVGNLTLE---AFGSEENQVVLGRD 377
DB 346 KNLDLIVANDVANPSIGFNSDENAITVIDRD 376
RESULT 7
ADP07150
ID ADF07150 standard; protein; 418 AA.
XX
XX ADF07150;
XX
XX 12-FEB-2004 (first entry)
XX
XX Bacterial polypeptide #3263.
XX

KW Proteus mirabilis infection; bacterial infection; antibacterial;
XX immunostimulant.
OS
XX Proteus mirabilis.
PN US6605709-B1.
XX 12-AUG-2003.
XX
XX 05-APR-2000; 2000US-00543681.
XX
XX 09-APR-1999; 99US-0128706P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL;
XX WPI; 2003-895291/82.
XX N-PSDB; ADF02978.
XX
XX New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX
XX Disclosure; SEQ ID NO 7435; 870pp; English.
XX
XX The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides, methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis,
CC method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX
SQ Sequence 418 AA;
Query Match 23.3%; Score 466; DB 7; Length 418;
Best Local Similarity 34.2%; Pred. No. 3.7e-36;
Matches 149; Conservative 76; Mismatches 145; Indels 66; Gaps 20;
QY 5 VKL---IYATK---SRKLVGKKIVXXPGSIAALDVKACEGLIRH---GAEVHAVMSEA 54
DB 3 VKLWTFIYAIRHFIMTTLHDKNLILGISGIAA--YKAPE-LVRLRDKGAIYRVVWTPA 59
QY 55 ATKIHPYAWNLTGNPVITEI-----TGFIEHVELAGEHENKADLIILVCPATANTISKI 109
DB 60 AHAFVTFLSLQAVSGFPFVADLLDPAAEAAMGHIELG---KWADLILLAPATADLIARL 115
QY 110 AGCIDTPTVTVTTFAPPHIPMIAPAMHETMYRHPVIRENIERLKKLVGEFIGP----- 164
DB 116 RMGMANDLLTTLCLAS--AAPAIAPAMNQMYRATITQENLTALAQRCGLIWGPOGSG 173
QY 165 ---RIEGR---AKVASIDEIVYRVIKLHKHKKLVGTAGATREYIDPIRFTTNA 217
DB 174 ACDVGPGRMLDPLELVALAEQOFAI-----QHDFFGKKITITAGPTREALDPVRFISNH 228
QY 218 SSGKMGVALAEADFRGA-VTLIR-----TKGSVKAFIRIKIKLVETVEEMLSAIENE 270
DB 229 SSGKMGFAIAQAAQAQGAETVLIAGPVTLPACVK-----RIDVESAQEMVQVMDI 281
QY 271 LRSKKYDVWVMAAAVSDFRPKKAESKI-KSGRSITIELVPXNPKIIDRI-KSIQPNVFL 328
DB 282 AQSQ--DIFGCAAVADYRAKQVAPKIKKQSGDEVITMI-KNPDIVASVGRKVEHRPFV 338
QY 329 VGFKAETSKELIECKRQRIERAKADLVVGNLTLE---AFGSEENQVVLGRDFTKELPK 384
DB 339 VGFAAET--QNVVEYARRKREKQKLDLICANDVSLDKDAGFNSDNLHLIHWANGETRLPH 396

```
QY 385 MKKRELAERIMDEIEK 400
Db 397 SSKTQLSHRLLEDEIAK 412

RESULT 8
ID AAG82741 standard; protein; 399 AA.
AC AAG82741;
XX
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis open reading frame protein sequence SEQ ID NO:2576.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
KW endocarditis.
XX
XX Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US030782.
XX
XX 09-NOV-1999; 99US-0164258P.
XX
XX (GLAX ) GLAXO GROUP LTD.
XX
XX Kimmerly WJ;
XX
XX WPI; 2001-316495/33.
DR N-PSDB; AAH53591.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.
XX
XX Claim 18; Page 675; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464
XX
XX Sequence 399 AA;
SQ
Query Match 23.1%; Score 461.5; DB 4; Length 399;
Best Local Similarity 34.3%; Pred. No. 9.5e-36;
Matches 141; Conservative 75; Mismatches 148; Indels 47; Gaps 16;
QY 19 KIVVXXPGSIALD-VKACEGLIRHGAHVHVMSEATKIHPYAWNLPTGNPVIT--- 74
Db 2 KHLIAVTGGIAAYKAIDTISKLIQSGYDVRVWMLSDHAQEFVTPLAQFASIRNPVYTNF 61
QY 75 --EITGFIEHVELAGEHENKADLILVCPATANTISKIACGI-DDTPVTVTVTATFPHIPI 131
```

```
Db 62 KEENPEIQHVSIG---DWADAIIVADATANTIAKLSVGIADDLITSTLLATTPK--- 114
QY 132 MIAPAMHETMYRHPIVRENIERLKLGVFEFIGP-----RIEGRKAVASIDE 178
Db 115 FVPAMNVMNPNRTHNMKVLSDGQYFFTEPGSGYLACGVAVAKRWE--PMQLSVIN 173
QY 179 IVYRVIKKHKKTLEGKRVLTAGATREYIDPIREITNASSGKMGVALAEADFRGA-VT 237
Db 174 KFFTQOKNVKSSPSGKALVTAGTVEVIDPVVSVNRSSGKMGYATAELRDKGAIIVT 233
QY 238 LIRTKGSVKAPIRIKIK-LKVETVEMLSALENELRSKKYDVIMAAVSDPRPKIKARG 296
Db 234 LI--SGPHTLSLPEGINVVKVESADDMFQAVTE--RFAKQDIVIKAAAVSDYTPMDILBH 289
QY 297 KIK---SGRSITIELVPXPKIIDRIKIQPNVLVGFKAETSKKLEEGKRQIERAKA 353
Db 290 KKKQEGSLSVQFK---RTKDILKYLGENKTHQYLVGFAAET--QNIQYALDKLKRKNA 344
QY 354 DLV---VGNLTLEAFGSENOVVLIGRDFTK-ELPKMKKRELAERIMDWIE 399
Db 345 DVIISNVGDTISIGFSSDDNLTWHFKNEKNKNIKKKKSALAHQIIELE 395

RESULT 9
ABP39940
ID ABP39940 standard; protein; 400 AA.
XX
XX AC ABP39940;
XX
XX 24-JUL-2002 (first entry)
XX
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4785.
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
XX Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-00134001.
XX
XX 14-AUG-1997; 97US-0055779P.
XX 08-NOV-1997; 97US-0064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2002-381255/41.
DR N-PSDB; ABN92485.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT polypeptide, useful for diagnosing and treating bacterial infections.
XX
XX Disclosure; SEQ ID NO 4785; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences can
CC also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life cycle
CC or inhibit S. epidermidis infection. N.B. The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site
XX
XX Sequence 400 AA;
```

Query Match 23.1%; Score 461.5; DB 5; Length 400;
Best Local Similarity 34.3%; Pred. No. 9.6e-36; Mismatches 148; Indels 47; Gaps 16;
Matches 141; Conservative 75; Mismatches 148; Indels 47; Gaps 16;

CC The invention describes an isolated nucleic acid comprising a nucleotide
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC 1-3772) and encoding an *Staphylococcus epidermidis* polypeptide with any
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC given in the specification. Also described are: a recombinant expression
CC vector; a cell comprising a recombinant expression vector of (1);
CC producing an *S. epidermidis* polypeptide; an isolated nucleic acid
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC vaccine composition for prevention or treatment of an *S. epidermidis*
CC infection, comprising a nucleic acid cited above and a carrier; treating
CC a subject for *S. epidermidis* infection; a recombinant or substantially
CC pure preparation of an *S. epidermidis* polypeptide or its fragment; a
CC vaccine composition for prevention or treatment of an *S. epidermidis*
CC infection; detecting the presence of a *Staphylococcus* nucleic acid in a
CC sample; a computer readable medium having recorded in it the nucleotide
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC system for identifying fragments of the *Staphylococcus* genome of
CC commercial importance; a computer based system for identifying fragments
CC of the *Staphylococcus* plasmids of commercial importance; identifying
CC commercially important nucleic acid fragments of the *Staphylococcus*
CC genome and/or plasmids; and identifying an expression modulating fragment
CC of the *Staphylococcus* genome and/or plasmids. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of an *Staphylococcus epidermidis* bacterial
CC infection. This is the amino acid sequence of a *S. epidermidis* protein of
CC the invention.
XX

QY 19 KKIIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAATKIIHPYAWNLPTGNPVIT--- 74
DB 3 KXILLAVTGGIAAYKAIDLTSKLIQSGYDVRVMSDHAQEFVTPLAFAQISRNPVYNTF 62

QY 75 --BITGFIEHVELAGHENKADILVCPATANTISKIACGI-DDTPVTVTVTTAPPHIPI 131
DB 63 KEENPEIQHVSIGL---DWADAIIVAPATANTIAKLSVGIADDLTSTLLATTPK--- 115

QY 132 MIPAMHETMYRHPVIRENTERLKLGVFEGP-----RIEGRKVASIDE 178
DB 116 FVAPANNVNNYNNPRTKHNKVLSDQGYFIEPGSYLACGYVAKGMEE-PMQILSVIN 174

QY 179 IVRVIRKHLKHTLEGKRVLTAGATREYIDPRTFNASSGKMGVALAEADFRGA-VT 237
DB 175 KPFTQKNVVKSSFSFGKRALVTAGPTVEVIDPVYVNSRSGKMGYALAEALDKGAIVT 234

QY 238 LIKTKGSVKAFIRKIK-LKVETVEMLSAIENELSKKYDVVIMAAVSDFRPKIKAE 296
DB 235 LI--SGPHTLSLPEGINVVKVESADDMFQAVTE--RFAKQDIVIKAAVSDYTPMDILEH 290

QY 297 KIK---SGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSEKELIEGKQIERAKA 353
DB 291 KLKQEGGLSVQPK---RTKDILKYLGENKTHQYLVGFAAET--QNIQYVALDKIKRKN 345

QY 354 DLV---VGNTEAFGSEENOVVLIGRDTFK-ELPKMKKRELAERIWEIDE 399
DB 346 DVIISNNVGDTSIGFSSDDNELTMHPKNEKVNKKGKKSALAHQIIEILE 396

RESULT 10
ADSO6229
ID ADSO6229 standard; protein; 400 AA.
XX
AC ADSO6229;
XX
DT 04-NOV-2004 (first entry)
XX
DE *Staphylococcus epidermidis* polypeptide seqid 5524.
XX
KW antibacterial; vaccine; antisense therapy; *Staphylococcus epidermidis*;
KW recombinant expression vector; infection; computer readable medium;
KW computer based system.
XX
OS *Staphylococcus epidermidis*.
XX
PN US2004147734-A1.
XX
PD 29-JUL-2004.
XX
PF 01-DEC-2003; 2003US-00724972.
XX
PR 08-NOV-1997; 97US-0064964P.
PR 13-AUG-1998; 98US-00134001.
PR 29-NOV-1999; 99US-00450969.
XX
XX (DOUC/) DOUCETTE-STAMM L.
PA (BUSH/) BUSH D.
XX
XX Doucette-Stamm L, Bush D;
XX
XX WPI: 2004-580138/56.
DR N-PSDB; ADS02457.
XX
XX New isolated polypeptide and encoding nucleic acid derived from
PT *Staphylococcus epidermidis*, useful for diagnosing, preventing and/or
PT treating an *S. epidermidis* bacterial infection.
XX
XX Claim 17; SEQ ID NO 5524; 741pp; English.
PS
XX

QY 19 KKIIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAATKIIHPYAWNLPTGNPVIT--- 74
DB 3 KXILLAVTGGIAAYKAIDLTSKLIQSGYDVRVMSDHAQEFVTPLAFAQISRNPVYNTF 62

QY 75 --BITGFIEHVELAGHENKADILVCPATANTISKIACGI-DDTPVTVTVTTAPPHIPI 131
DB 63 KEENPEIQHVSIGL---DWADAIIVAPATANTIAKLSVGIADDLTSTLLATTPK--- 115

QY 132 MIPAMHETMYRHPVIRENTERLKLGVFEGP-----RIEGRKVASIDE 178
DB 116 FVAPANNVNNYNNPRTKHNKVLSDQGYFIEPGSYLACGYVAKGMEE-PMQILSVIN 174

QY 179 IVRVIRKHLKHTLEGKRVLTAGATREYIDPRTFNASSGKMGVALAEADFRGA-VT 237
DB 175 KPFTQKNVVKSSFSFGKRALVTAGPTVEVIDPVYVNSRSGKMGYALAEALDKGAIVT 234

QY 238 LIKTKGSVKAFIRKIK-LKVETVEMLSAIENELSKKYDVVIMAAVSDFRPKIKAE 296
DB 235 LI--SGPHTLSLPEGINVVKVESADDMFQAVTE--RFAKQDIVIKAAVSDYTPMDILEH 290

QY 297 KIK---SGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSEKELIEGKQIERAKA 353
DB 291 KLKQEGGLSVQPK---RTKDILKYLGENKTHQYLVGFAAET--QNIQYVALDKIKRKN 345

QY 354 DLV---VGNTEAFGSEENOVVLIGRDTFK-ELPKMKKRELAERIWEIDE 399
DB 346 DVIISNNVGDTSIGFSSDDNELTMHPKNEKVNKKGKKSALAHQIIEILE 396

Query Match 23.1%; Score 461.5; DB 8; Length 400;
Best Local Similarity 34.3%; Pred. No. 9.6e-36; Mismatches 148; Indels 47; Gaps 16;
Matches 141; Conservative 75; Mismatches 148; Indels 47; Gaps 16;

CC The invention describes an isolated nucleic acid comprising a nucleotide
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC 1-3772) and encoding an *Staphylococcus epidermidis* polypeptide with any
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC given in the specification. Also described are: a recombinant expression
CC vector; a cell comprising a recombinant expression vector of (1);
CC producing an *S. epidermidis* polypeptide; an isolated nucleic acid
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC vaccine composition for prevention or treatment of an *S. epidermidis*
CC infection, comprising a nucleic acid cited above and a carrier; treating
CC a subject for *S. epidermidis* infection; a recombinant or substantially
CC pure preparation of an *S. epidermidis* polypeptide or its fragment; a
CC vaccine composition for prevention or treatment of an *S. epidermidis*
CC infection; detecting the presence of a *Staphylococcus* nucleic acid in a
CC sample; a computer readable medium having recorded in it the nucleotide
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC system for identifying fragments of the *Staphylococcus* genome of
CC commercial importance; a computer based system for identifying fragments
CC of the *Staphylococcus* plasmids of commercial importance; identifying
CC commercially important nucleic acid fragments of the *Staphylococcus*
CC genome and/or plasmids; and identifying an expression modulating fragment
CC of the *Staphylococcus* genome and/or plasmids. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of an *Staphylococcus epidermidis* bacterial
CC infection. This is the amino acid sequence of a *S. epidermidis* protein of
CC the invention.
XX

QY 19 KKIIVXXPGSIAALD-VKACEGLIRHGAHVHVMSEAATKIIHPYAWNLPTGNPVIT--- 74
DB 3 KXILLAVTGGIAAYKAIDLTSKLIQSGYDVRVMSDHAQEFVTPLAFAQISRNPVYNTF 62

QY 75 --BITGFIEHVELAGHENKADILVCPATANTISKIACGI-DDTPVTVTVTTAPPHIPI 131
DB 63 KEENPEIQHVSIGL---DWADAIIVAPATANTIAKLSVGIADDLTSTLLATTPK--- 115

QY 132 MIPAMHETMYRHPVIRENTERLKLGVFEGP-----RIEGRKVASIDE 178
DB 116 FVAPANNVNNYNNPRTKHNKVLSDQGYFIEPGSYLACGYVAKGMEE-PMQILSVIN 174

QY 179 IVRVIRKHLKHTLEGKRVLTAGATREYIDPRTFNASSGKMGVALAEADFRGA-VT 237
DB 175 KPFTQKNVVKSSFSFGKRALVTAGPTVEVIDPVYVNSRSGKMGYALAEALDKGAIVT 234

QY 238 LIKTKGSVKAFIRKIK-LKVETVEMLSAIENELSKKYDVVIMAAVSDFRPKIKAE 296
DB 235 LI--SGPHTLSLPEGINVVKVESADDMFQAVTE--RFAKQDIVIKAAVSDYTPMDILEH 290

QY 297 KIK---SGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSEKELIEGKQIERAKA 353
DB 291 KLKQEGGLSVQPK---RTKDILKYLGENKTHQYLVGFAAET--QNIQYVALDKIKRKN 345

QY 354 DLV---VGNTEAFGSEENOVVLIGRDTFK-ELPKMKKRELAERIWEIDE 399
DB 346 DVIISNNVGDTSIGFSSDDNELTMHPKNEKVNKKGKKSALAHQIIEILE 396

RESULT 11
ADL05850
ID ADL05850 standard; protein; 431 AA.
XX
AC ADL05850;
XX
DT 06-MAY-2004 (first entry)
XX
DE M. catarrhalis protein #1616.
XX
XX Moraxella catarrhalis; infection.

```
XX OS Moraxella catarrhalis.
XX PN US6673910-B1.
XX PD 06-JAN-2004.
XX PF 04-APR-2000; 2000US-00540236.
XX PR 08-APR-1999; 99US-0128416P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Breton GL;
XX WPI; 2004-178127/17.
XX DR N-PSDB; ADL03930.
XX PT New nucleic acid encoding a Moraxella catarrhalis polypeptide, useful for
XX PT preparing a composition for diagnosing, preventing or treating infection
XX PT caused by Moraxella catarrhalis.
XX PS Disclosure; SEQ ID NO 3536; 429pp; English.
XX CC The invention relates to an isolated nucleic acid encoding an Moraxella
XX CC catarrhalis polypeptide. The nucleic acid is useful for preparing a
XX CC composition for diagnosing, preventing or treating infection caused by
XX CC Moraxella catarrhalis. The present sequence represents the amino acid
XX CC sequence of a M. catarrhalis protein.
XX SQ Sequence 431 AA;
Query Match 22.9%; Score 458.5; DB 8; Length 431;
Best Local Similarity 33.0%; Pred. No. 2.1e-35;
Matches 140; Conservative 88; Mismatches 147; Indels 49; Gaps 19;
QY 5 VKLIVATSKRLVGK--KIVXXXPGSIAAL-DVKACEGLIRHGAHVAMVSEAATKIHP 61
DB 11 IMNVFAKCMQVNTQPKILLAVTGIIAAYKSAMLARLLIKSGCQVRVMTAGACEFITP 70
QY 62 YAWNLPNGPVITEITG-----FIEHVELAGEHENKADLILVCPATANTISKIACGIDDT 116
DB 71 LTFQALTGOAVHTKLLDETAERGGMGHITLA---KWADLVVVAPASANTIAKLAGGFADN 126
QY 117 PVTVVTTAFPHPIAMIPAMHETMYRHPVIVRENIERLKKLGVETFGPRIE-----GR 170
DB 127 LVTTVCLAT--DAPILIAPANMNMQWANAIVQDNLKLRQFGYHIMTPDSGQACGDVGL 184
QY 171 AKVASIDEIVYRVI---KKLH-KKLEGRKRVLTAGATREYIDPIRFTNASSGKMGVAL 226
DB 185 GRLEPEDLCQOILAFDKCRHLOPSLLGKTVITAGATVEPIDPVRFLSNHSTGKMGVAL 244
QY 227 ABEADFRGA-VTLIRTKG-SVKA-FRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAA 283
DB 245 ANACYHAGADVILVSGKHVSLKTPFGVR--RLDVGTAEOMLKVCLSV--CQKADIFIAA 300
QY 284 AVSDRPPKIKAGSKI---KSGRSITTELVPXNPKIIDRIKEIQPNVFLVGFPAET----- 335
DB 301 AVADFKAAASVADHKIKTKNQOTMTLQLI-KNPDVLATICTHTYPDLICVGFPAETQDANN 359
QY 336 -SKEKLEEGKQIERAKADLVVGNLTLEAFGSEENOVVLIGRD---FTKELPKMKKREL 390
DB 360 CAKSJL---AAKQOLDMIANDVSDKTI-GFGSENNAMTVFFAEQYDQMPQLPKAHKSKI 415
QY 391 AERI 394
DB 416 AEQL 419
RESULT 12
ABP80942
ID ABP80942 standard; protein; 394 AA.
XX
```

```
AC ABP80942;
XX 07-MAR-2003 (first entry)
XX N. gonorrhoeae amino acid sequence SEQ ID 8414.
XX DE Antibacterial; infection; vaccine; gene therapy.
XX KW Neisseria gonorrhoeae.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB002069.
XX PR 12-FEB-2001; 2001GB-00003424.
XX PA (CHIR-) CHIRON SPA.
XX PI Fontana MR, Pizza M, Masignani V, Monaci E;
XX WPI; 2003-058415/05.
XX DR N-PSDB; ABZ41912.
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX PT medicament for treating or preventing N. gonorrhoeae infection.
XX PS Disclosure; Page 801; 815pp; English.
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
XX CC antibodies that specifically bind to the proteins. The composition
XX CC comprising the protein, nucleic acid or antibody is useful for the
XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae
XX CC infection, this may be in the form of a vaccine or gene therapy.
XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX CC molecules of the invention
XX SQ Sequence 394 AA;
Query Match 22.9%; Score 457; DB 6; Length 394;
Best Local Similarity 33.7%; Pred. No. 2.6e-35;
Matches 138; Conservative 72; Mismatches 161; Indels 38; Gaps 15;
QY 17 VKKIVXXXPGSIAALDVKACEGLIR---HCAEVHVMSEAATKIHPYANLPTGNPV 72
DB 1 MKGHILLGVGTGIAA--YKSCE-LVRLKKQGHSTVTVMSRSATFVSLPTQALSGNPV 57
QY 73 ITEITG-----FIEHVELAGEHENKADLILVCPATANTISKIACGIDDTPTVTVVTTAPP 127
DB 58 LTDTHGGNGSGMEHINT---RNADVFLIAPASMTVAKICNGVADNLLTNL--AAAR 111
QY 128 HIPIMIAPIAMHETMYRHPVIVRENIERLKKLGVETFGPRI-----EGRKAVASIDEIVY 181
DB 112 KCPLAIPAMNVMMNLNPNQRIANQLVSDGITVVMPLGEGQACRENGMGRPEPAELLD 171
QY 182 RVIKKLHKKTLEGRVLTAGATREYIDPIRFTNASSGKMGVALAEEADFRGA-VTLIR 240
DB 172 LLPDLWTTPKILRKKILITAGATFEAIDPVRGITTSSGKMGVALARACRAAGAEISLH 231
QY 241 TKGSKVAFRIKIKLVETV--EEMLSAIEENELRSKKYDVVIMAAVSDFRPKIAEGKI 298
DB 232 --GQLOTTLPFGISITVQAVSAEDMHRVHRLI--EKQDAFISVAASVDYKVKNRSTQKF 287
QY 299 ---KSGRSITTELVPXNPKIIDRIKEIQPNVFLVGFKAFTSKEKLEEGKQIERAKADL 355
DB 288 KKDKNAKPLSIEL-DENPDTLASIASLPNPPFCIGFAAET--ENVMAVAREKRIKKKIPV 344
QY 356 VVGNTLE-AFGSEENOVVLIGRDFTKELPKMKKRELARIWDEIEKXLS 403
DB 345 IVANDVSIAMGKTTNQIIVIDDDAELSPFETSDEAAMRIVERLAVYLN 393
```



```
RESULT 13
ABM71249
ID ABM71249 standard; protein; 401 AA.
XX
AC ABM71249;
XX
DT 20-NOV-2003 (first entry)
XX
DE Staphylococcus aureus protein #489.
XX
KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;
KW enzymatic assay; antibiotic target.
XX
OS Staphylococcus aureus.
XX
PN WO200294868-A2.
XX
PD 28-NOV-2002.
XX
PF 27-MAR-2002; 2002WO-IB002637.
XX
PR 27-MAR-2001; 2001GB-00007661.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Masignani V, Mora M, Scarselli M;
XX
DR WPI; 2003-120786/11.
DR N-PSDB; ACF72809.
XX
PT New Staphylococcus aureus protein, useful as a vaccine for treating or
PT preventing Staphylococcal infection, specifically an infection caused by
PT S. aureus, e.g. sepsis.
XX
PS Claim 1; SEQ ID NO 978; 49pp; English.
XX
CC The invention relates to novel genes and encoded proteins from
CC Staphylococcus aureus. A composition comprising the S. aureus protein, a
CC nucleic acid encoding the protein, or an antibody to the protein, is
CC useful as a pharmaceutical, particularly as a vaccine for treating or
CC preventing infection due to Staphylococcus bacteria, specifically an
CC infection caused by S. aureus. The composition is particularly useful for
CC treating or preventing sepsis in a patient. The composition can also be
CC used for diagnostics. The protein is also used in an assay for enzymatic
CC studies and as a target for antibiotics. This sequence represents one of
CC the novel S. aureus proteins of the invention
XX
SQ Sequence 401 AA;

Query Match      22.7%; Score 453.5; DB 6; Length 401;
Best Local Similarity 32.7%; Pred. No. 5.8e-35;
Matches 133; Conservative 81; Mismatches 154; Indels 39; Gaps 13;

QY 19 KIVXXXPSGIAALD-VKACEGLIRHGAENVHVMSEAATKIHPYAMNLPNGNVT--- 74
Db 4 KXILLAVTGIAAYKAIDLTKLTQSGYEVVRVMTLNHAQKFVTPLAFOAISRNAVYDTDF 63
QY 75 --EITGFIHVELAGHENKADLLVCPATANTISKIACGI-DDTPVTVTVTAPPHIPI 131
Db 64 IENPSEIOHIALG-----DWADAIIVAPATANTIAKLSVGIADDLVTSTLLAT---ETPK 116
QY 132 MIAPAMHETMYRHPIVRENIERLKKLGVEFIGP-----RIEGRKAVASIDE 178
Db 117 FIAPAMNVMHYENKRTQONINILKEDGYHFIPEGSGFLACGYVAKGRMEE-PLQIVSVID 175
QY 179 IVYRVIKKLHKKTLEGRVLTAGATREYIDPIRFTFNASSGKMGVALAEEDFRAQ-VT 237
Db 176 AHFQNSNRLANSFQDKRALVTAGPTIEVDPRFVSNRSSGKMGVIAEALRNRAIVT 235
QY 238 LIRTKGSVKAIRIKLKVETVEEMLSAIENELRSKDYVIMAAVSDPRPKIAEGK 297
Db 236 LVAGPTTLEDPKDIEV-IHVQSAEENFEQVTS--RFEQDQIVVKAADVSDYTFVDVLEHK 292

QY 298 IKSGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKKELIBGKQIERAKADLVV 357
Db 293 MKKQDGLSVSPKTKDILKYLGEHKTSOYLIGFAET--EDIENYAOQKLRKNKADVII 350
QY 358 GNTLE---AFGSEENQVVLIGRDTFK-ELPKMKKRELAERIWDTE 399
Db 351 SNNVGDMSIGFSSDDNLTMHFNKEKVNKKKKVVLAAQILDELE 397

RESULT 14
ADB10278
ID ADB10278 standard; protein; 401 AA.
XX
AC ADB10278;
XX
DT 20-NOV-2003 (first entry)
XX
DE Alloiooccus otitis antigenic protein SEQ ID NO:4782.
XX
KW Alloiooccus otitis; antigenic protein; immunogenic; immunisation;
KW gene therapy; Gram-positive bacterium; infection.
XX
OS Alloiooccus otitis.
XX
PN WO2003048304-A2.
XX
PD 12-JUN-2003.
XX
PF 25-NOV-2002; 2002WO-US036123.
XX
PR 29-NOV-2001; 2001US-0333777P.
PR 18-NOV-2002; 2002US-0426742P.
XX
PA (AMHP ) WYETH HOLDINGS CORP.
XX
PI Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;
XX
DR WPI; 2003-505284/47.
DR N-PSDB; ADB10281.
XX
CC New Alloiooccus otitis polynucleotides and polypeptides, useful for
CC treating and diagnosing diseases, drug screening assays and monitoring of
CC effects during drug clinical trials.
XX
PS Claim 33; SEQ ID NO 4782; 1019pp; English.
XX
CC The present invention describes an isolated polynucleotide (I) of
CC Alloiooccus otitis genomic DNA, which encodes an antigenic protein.
CC Alloiooccus otitis is a Gram-positive bacterium. Also described: (1)
CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
CC expression vector comprising the novel isolated polynucleotide (I), its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC of the polypeptides of (1), their biological equivalent or fragment; (8)
CC immunising against Alloiooccus otitis by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying Alloiooccus
CC otitis in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (I) can be used in gene therapy. The
CC polynucleotides, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting Alloiooccus
CC otitis. The present sequence represents an Alloiooccus otitis
CC antigen protein from the present invention.
```

XX SQ Sequence 401 AA;
Query Match 22.2%; Score 444; DB 6; Length 401;
Best Local Similarity 32.7%; Pred. NO. 4.9e-34;
Matches 137; Conservative 81; Mismatches 145; Indels 56; Gaps 18;
QY 16 LVGKIVXXPGSIAL--DVKACEGLIRHGAHVHVMSEAATKIHPVAMNLTGPNVIT 74
DB 2 LKNKIALYVTGGIAVYKSLYLLREIIKQGEVRVMTQAACQFVNPISFQVLSQKVV-- 59
QY 75 EITGF-----IEHVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTAPP 127
DB 60 QIDTFEEGQPEVSVIDLT---DWADYSIVAPATANIIGKLANGIGDDFVSTAL-LATD 114
QY 128 HIPMIAPAMHETMYRHPVIRENTERLKKLGVEFGIPRI-----EGRKVASIDEIV- 180
DB 115 H-PIPLVPAMNTKMYENPALKKNKAFLIEQGHYMWEPDIFGLAEGYGLGRFPDLDRIMA 173
QY 181 ---YRVIKLHKHTLEGGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-V 236
DB 174 EFNHFIAR-NPGILSGKKVLVTAGTVERIDPVRIYISNDSSGKMGHQLAQAAEYAGQV 232
QY 237 TLIR-----TKGSVKAFRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPK 291
DB 233 SLVTASDLPTSPFIDRFQVESTLDLYQTVSDLYD-----HHDILMMAAAVSDYRPV 283
QY 292 IKAEGKIKSGRSITIELVPXNPKII---DRIKEIQPNVFLVGFKAETSKELIEGKROI 348
DB 284 NRSDDKMKKQDNLTIEL-EKNPDILAEMGRKKDQIN---VGFAAET--HNLEEYAKKL 337
QY 349 ERAKADLVVGNLT---EAFGSEENQVVLIGRDFTK-ELPKMKKRELAERIWDIEKKL 402
DB 338 ASKQADLIVANEVGRDGRGNADENALVSSDQDPLELPLOSKKDMAKKIIIEVVASKL 396

RESULT 15
ADJ27133 standard; protein; 401 AA.
XX AC ADJ27133;
XX DT 20-MAY-2004 (first entry)
XX DE Alloicoccus otitidis protein, a novel antibacterial target SeqID 78.
XX KW growth and survival; otitis media with effusion; OME;
XX KW bacterial infection; antibacterial; antiinfective.
XX OS Alloicoccus otitis.
XX FN WO2003104391-A2.
XX PD 18-DEC-2003.
XX PF 25-NOV-2002; 2002WO-US036122.
XX PR 29-NOV-2001; 2001US-0333777P.
XX PR 18-NOV-2002; 2002US-0426742P.
XX PR 18-NOV-2002; 2002US-0427367P.
XX XX (AMHP) WYETH HOLDINGS CORP.
XX PI Murphy E, Projan SJ;
XX DR WPI; 2004-053616/05.
XX DR N-PSDB; ADJ27132.
XX PT Novel isolated Alloicoccus otitidis polypeptides useful for identifying
XX PT compounds that inhibit the activity or expression of the polypeptide and
XX PT thus are useful for treating infections caused by Alloicoccus otitidis.
XX PS Claim 7; SEQ ID NO 78; 433pp; English.

XX CC This invention relates to novel isolated Alloicoccus otitidis (A.
CC otitidis) nucleic acid molecules and encoded proteins thereof.
CC Specifically, it refers to proteins that are essential for the growth and
CC survival of the gram-positive bacterium A. otitidis, and hence provide
CC novel antibacterial targets. The present invention describes
CC pharmaceutical compositions and antisense compounds that are useful for
CC inhibiting activity or expression of these proteins. Furthermore, it
CC provides diagnostics and therapeutics that can be used to ameliorate
CC diseases that are associated with A. otitidis, such as otitis media with
CC effusion (OME) and various bacterial infections. Accordingly, these
CC compositions exhibit both antibacterial and antiinfective activities.
CC This polypeptide is an A. otitidis protein sequence of the invention.
XX SQ Sequence 401 AA;
Query Match 22.2%; Score 444; DB 8; Length 401;
Best Local Similarity 32.7%; Pred. NO. 4.9e-34;
Matches 137; Conservative 81; Mismatches 145; Indels 56; Gaps 18;
QY 16 LVGKIVXXPGSIAL--DVKACEGLIRHGAHVHVMSEAATKIHPVAMNLTGPNVIT 74
DB 2 LKNKIALYVTGGIAVYKSLYLLREIIKQGEVRVMTQAACQFVNPISFQVLSQKVV-- 59
QY 75 EITGF-----IEHVELAGEHENKADLILVCPATANTISKIACGIDTPTVTTVTAPP 127
DB 60 QIDTFEEGQPEVSVIDLT---DWADYSIVAPATANIIGKLANGIGDDFVSTAL-LATD 114
QY 128 HIPMIAPAMHETMYRHPVIRENTERLKKLGVEFGIPRI-----EGRKVASIDEIV- 180
DB 115 H-PIPLVPAMNTKMYENPALKKNKAFLIEQGHYMWEPDIFGLAEGYGLGRFPDLDRIMA 173
QY 181 ---YRVIKLHKHTLEGGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-V 236
DB 174 EFNHFIAR-NPGILSGKKVLVTAGTVERIDPVRIYISNDSSGKMGHQLAQAAEYAGQV 232
QY 237 TLIR-----TKGSVKAFRIRKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPK 291
DB 233 SLVTASDLPTSPFIDRFQVESTLDLYQTVSDLYD-----HHDILMMAAAVSDYRPV 283
QY 292 IKAEGKIKSGRSITIELVPXNPKII---DRIKEIQPNVFLVGFKAETSKELIEGKROI 348
DB 284 NRSDDKMKKQDNLTIEL-EKNPDILAEMGRKKDQIN---VGFAAET--HNLEEYAKKL 337
QY 349 ERAKADLVVGNLT---EAFGSEENQVVLIGRDFTK-ELPKMKKRELAERIWDIEKKL 402
DB 338 ASKQADLIVANEVGRDGRGNADENALVSSDQDPLELPLOSKKDMAKKIIIEVVASKL 396

Search completed: July 2, 2005, 12:22:41
Job time : 89.0698 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 2, 2005, 08:53:17 ; Search time 2821 Seconds
(without alignments)
2104.937 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPDWKRKEILIEPSEEE.....PYRGNYQSGTRIAFSKRKL 156

Scoring table: BLOSUM62
-Q=/cgn2 1/USPTO spool/US08957709/runat 01072005 154632 26016/app query.fasta_1.327
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US08957709 @CGN 1 1 4352 @runat 01072005 154632 26016 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2 1/USPTO spool/US08957709/runat 01072005 154632 26016/app query.fasta_1.327
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US08957709 @CGN 1 1 4352 @runat 01072005 154632 26016 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gsa1:
9: gb_gsa2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	23.8	600	BZ893227	HL12_0182
2	162.5	20.2	792	CL680050	CL680050 PRI0127C
3	150	18.6	1089	BZ551941	BZ551941 pac61-60
4	125	15.5	465	CD084073	CD084073 MA3-9999SU
5	124	15.4	586	CB336913	CB336913 Tc024E08F
6	122	15.1	491	A1395397	A1395397 MA003489
7	122	15.1	507	CD145288	CD145288 MGI-0104U
8	122	15.1	530	CB335318	CB335318 Tc005E10R
9	119	14.8	800	CK315881	CK315881 SB02028A1

10	119	14.8	826	7	CK311070
11	118.5	14.7	519	1	AA968575
12	118.5	14.7	839	8	BZ549391
13	117	14.5	663	1	AJ451538
14	116.5	14.5	481	7	CO297884
15	116	14.4	444	7	CV037326
16	116	14.4	474	6	CB018433
17	116	14.4	519	6	CB110414
18	116	14.4	587	5	BX336627
19	116	14.4	597	5	BX262764
20	116	14.4	597	5	BX262765
21	116	14.4	606	5	BX304266
22	116	14.4	608	1	AJ445450
23	116	14.4	632	6	CD216556
24	116	14.4	636	5	BUI04949
25	116	14.4	642	4	BM486691
26	116	14.4	648	5	BX395616
27	116	14.4	653	5	BX390007
28	116	14.4	664	6	CD218033
29	116	14.4	676	5	BUI12250
30	116	14.4	686	5	BUI296179
31	116	14.4	701	7	CO422270
32	116	14.4	714	6	CD762467
33	116	14.4	801	5	BUI421023
34	116	14.4	829	5	BUI13346
35	116	14.4	830	5	BUI27586
36	116	14.4	842	5	BUI17711
37	116	14.4	925	5	BUI08606
38	116	14.4	937	7	CO648549
39	116	14.4	1066	5	BUI473281
40	115.5	14.3	521	4	B1236417
41	115.5	14.3	522	3	AY231831
42	115.5	14.3	523	4	B1233284
43	115.5	14.3	534	4	B1236440
44	115.5	14.3	545	1	A1512510
45	115.5	14.3	548	4	B1173237

ALIGNMENTS

RESULT 1
BZ893227
LOCUS HL12_0182 H1 pUC18 Library Halorubrum lacusprofundi genomic 5',
DEFINITION genomic survey sequence.
ACCESSION BZ893227
VERSION BZ893227.1
KEYWORDS GI:33343817
SOURCE GSS.
ORGANISM Halorubrum lacusprofundi
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halorubrum.
REFERENCE 1 (bases 1 to 600)
AUTHORS Goo,Y., Roach,J., Glusman,G., Baliga,N.S., Deutsch,K., Pan,M., DasSarma,S., Ng,W.V. and Hood,L.
TITLE Low-pass Sequencing for Microbial Comparative Genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Goo Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoo@systemsbiology.org
Seq primer: M13 Forward
Class: shotgun
Location/Qualifiers
1..600
/organism="Halorubrum lacusprofundi"
/mol_type="genomic DNA"
/strain="ATCC 49239"
/db_xref="taxon:2247"
/clone_lib="H1 pUC18 Library"
/note="Vector: pUC18; Site_1: SmaI; A shotgun library was

constructed from Halorubrum lacusprofundi genomic DNA
using pUC19/Snai/BAP plasmid"

ORIGIN

Alignment Scores:
Pred. No.: 5e-14 Length: 600
Score: 192.00 Matches: 55
Percent Similarity: 50.57% Conservative: 34
Best Local Similarity: 31.25% Mismatches: 53
Query Match: 23.82% Indels: 34
DB: 8 Gaps: 8

US-08-957-709A-71 (1-156) x BZ893227 (1-600)

```
Qy 1 MetLeuLeuProAspTyrLysIle-----ArgLysGluLeuLeuGlu 15
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 31 ATGATCTCTCGGACGCGGACATCTCGACCGCTCGCGGAGGGGACCTCGTCGTGAA 90
Qy 16 PropheSer-----GluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 91 CCCCTCGACGACGTCGACACGAGGTCCAGCCCGGAGCGTCGACCTCGGACTCGGC--- 147
Qy 34 GluAlaPheValLys-----Gly 39
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 148 GAGCGCTTCTCGGAGTTCGAGCGCTTATCATCTCCATCCACCCAAACGAGCGCGAC 207
Qy 40 LysLeuIleAspValGluLysGluLysVal-----ValIlePro 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 208 GAGTCTCGCGACTACGTCCAGGACGCGCTCGCGGAGCGGAGGATTCATCTCTCCAC 267
Qy 54 ProArgGluTyrAlaLeuLeuLeuThruLeuGluArgLysLeuProAspValMet 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 268 CCGGCGGACTCGCTCGGACACACGAGCGCGCTCGAGATCCCGCGGACTGCTC 327
Qy 74 GlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIle-----GlySerPhe 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 328 GCGACCGGTGAAGCGCGCTCTCGCTCGGGGGCTCGCGACTCATCCAGCGCCACCGCG 387
Qy 92 AlaTrpValAspProGlyTyrAspGlyAsnLeuThruLeuMetLeuTyrAsnAlaSerAsn 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 388 GGGATCGTCGATCCGGGGTACAAGGGGCGAGATCGACTCGAACTTCGAACTTCGGAC 447
Qy 112 GluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGlu 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 448 GNGCCGTCGGCTCACCGCGGATGCGCTCTCACAGCTCATCTTCACGGAGCTCAA 507
Qy 132 GlyProAlaArgAsnProTyr-----ArgGlyAsn-----TyrGln 143
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 508 TCGCCCGCAAAACGCCCGCTATGGCGTCGACGCGGCTCGAAGTATCAG 555
```

RESULT 2

CL680050/c 792 bp DNA linear GSS 09-JUL-2004
LOCUS PRI0127c.H08.2 - PRI0127c.BR (792) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.

ACCESSION CL680050
VERSION CL680050.1 GI:50186880
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 792)
AUTHORS Srinivasan, J., Otto, G.W., Kahlow, U., Geisler, R. and Sommer, R.J.
TITLE AppADB: an AcedB database for the nematode satellite organism
Pristionchus pacificus
JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel.: 00497071601371

Fax: 00497071601498
Email: ralf.sommer@uebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.
Location/Qualifiers
1. 792
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

FEATURES
source

ORIGIN

Alignment Scores:
Pred. No.: 4.69e-10 Length: 792
Score: 162.50 Matches: 41
Percent Similarity: 50.93% Conservative: 14
Best Local Similarity: 37.96% Mismatches: 44
Query Match: 20.16% Indels: 9
DB: 9 Gaps: 2

US-08-957-709A-71 (1-156) x CL680050 (1-792)

```
Qy 54 ProArgGluTyrAlaLeuLeuThruLeuGluArgLysLeuProAspValMet 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 672 CCAGGAGAGTGGCGGTGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 613
Qy 74 GlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla--- 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 612 GGCTGGCTGACGCGCGCTTCTCTTTGGCGGTCTGGGGCTGATGGTGCACGTCAC 553
Qy 93 ---TrpValAspProGlyTyrAspGlyAsnLeuThruLeuMetLeuTyrAsnAlaSerAsn 111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 552 CACCGTATCATCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 493
Qy 112 GluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGlu 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 492 CTGCGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 433
Qy 132 GlyProAlaArgAsnProTyrArg-----GlyAsnTyrGlnGly 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 432 GGCGCGGCGCGCGCTTACACCGCGGTGAGATGCGAAATATCGCAACACGAGG 373
Qy 145 SerThrArgLeuAlaPheSerLys 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 372 GCGGTAGCCAGCGCGATCGATATA 349
```

RESULT 3

BZ551941 1089 bp DNA linear GSS 17-DEC-2002
LOCUS pacsl-60.3403.x1 pacsl-60 Pseudomonas aeruginosa genomic clone
pacsl-60.3403, genomic survey sequence.

ACCESSION BZ551941
VERSION BZ551941.1 GI:27155771
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1089)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA

LOCUS CB336913 586 bp mRNA linear EST 01-JAN-2004
DEFINITION Tc024E08F Tribolium castaneum embryonic cDNA library Tribolium
castaneum cDNA clone Tc024E08 3', mRNA sequence.

ACCESSION CB336913
VERSION CB336913.1 GI:40544638

SOURCE EST.
ORGANISM Tribolium castaneum (red flour beetle)

Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
Tenebrionidae; Tribolium.

REFERENCE 1 (bases 1 to 586)

AUTHORS Savard, J. and Tautz, D.

TITLE A Tribolium castaneum EST project

JOURNAL Unpublished (2003)

COMMENT Contact: Savard, J.

Abteilung fur Evolutionsgenetik, AG Tautz

Institut fur Genetik, Universitat zu Koln

Weyertal 121, 50931 Koln, Germany

Tel: 49 221 470 6911

Fax: 49 221 470 5975

Email: savard@uni-koeln.de

Seq primer: M13F -20.

Location/Qualifiers

1..586

/organism="Tribolium castaneum"

/mol_type="mRNA"

/strain="wild type"

/db_xref="taxon:7070"

/clone="Tc024E08"

/dev_stage="Mixed embryonic stages"

/clone_lib="Tribolium castaneum embryonic cDNA library"

/note="Vector: pBluescript SK; Site 1: EcoRI; Site 2:

XhoI; Uni-ZAP XR cDNA library (Stratagene) constructed by

Reinhard Schroder (1995)"

ORIGIN

Alignment Scores:
Pred. No.: 2,85e-05 Length: 586
Score: 124.00 Matches: 43
Percent Similarity: 50.38% Conservative: 24
Best Local Similarity: 32.33% Mismatches: 43
Query Match: 15.38% Indels: 24
DB: 6 Gaps: 4

US-08-957-709A-71 (1-156) x CB336913 (1-586)

Qy 16 PropheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAla 35

Db 496 CCTCGACCAAGGTTTCAGTTAAAGCCGCGGTATGATCTG-----AAGAGCGCT 446

Qy 36 PheValLysGlyLysLeuLeuAspValGluLysGluGlyLysValValleProProArg 55

Db 445 TTT-----GACTGCGTGGTCCCGCCGA 422

Qy 56 GluTyrAlaLeuLeuThrLeuGluArgIleLysLeuProAspValMetGlyAsp 75

Db 421 GGTAAAGGCCCTCGTGGACCCGCGCATCAAAATCCAGCTCCCGAAGGGTGTCAGGCGG 362

Qy 76 MetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaTrp---Val 94

Db 361 ATCGCCCTAGTCTGGGCTAGCAGTCAAGAAATTCATCGATGTGGCGCGGGGTCGTG 302

Qy 95 AppProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProVal 114

Db 301 GATGAGGACTACAGGGGGGTCTCTAAGGTGGTCTGTTCATCCATTACAGTACTCCGTTT 242

Qy 115 GluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAla 134

Db 241 GAAGTCAAGAGTGGCGCCGATTTCACAGCTAAATTCGCGCGCGATT----- 194

Qy 135 ArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArg 147

Db 193 -----TATTA-CCCCGATATCGAGGAAGTTCAAGA 165

RESULT 6

LOCUS AI395397/c

DEFINITION MA003489.C8F Soares normalized S8W Schistosoma mansoni cDNA 3',
mRNA sequence.

ACCESSION AI395397

VERSION AI395397.1 GI:4224944

KEYWORDS EST.

SOURCE Schistosoma mansoni

ORGANISM Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

REFERENCE 1 (bases 1 to 491)

AUTHORS Bailey, J.A., Bentley, K., Blanton, R.E., Soares, M.B. and

Chakravarti, A.

TITLE Expressed sequence tags from a normalized adult Schistosoma mansoni

library

JOURNAL Unpublished (1999)

COMMENT Contact: Chakravarti, A.

Department of Genetics

Case Western Reserve University

2109 Adelbert Rd, Cleveland, OH 44106, USA

Tel: 216-368-5847

Fax: 216-368-5857

Email: axc39@po.cwru.edu

Additional data regarding this EST may be found at

http://genome.cwru.edu/schistosoma/est/S8west.html

Seq primer: M13 Forward.

Location/Qualifiers

1..491

/organism="Schistosoma mansoni"

/mol_type="mRNA"

/strain="Puerto Rican"

/db_xref="taxon:6183"

/sex="mixed"

/dev_stage="8 week old adult worms"

/lab_host="Mus musculus"

/clone_lib="Soares normalized S8W"

/note="Vector: p773-pac; Site_1: NotI; Site_2: EcoRI;

Note: normalization has resulted in an increased frequency

of the rare contaminating mouse (host) sequences (~10%).

For details of library construction see: Bonaldo MF,

Lennon G, Soares MB. Normalization and Subtraction: Two

Approaches to Facilitate Gene Discovery. 1996. Genome

Research 6:791-806"

ORIGIN

Alignment Scores:

Pred. No.: 4.03e-05 Length: 491

Score: 122.00 Matches: 41

Percent Similarity: 42.96% Conservative: 17

Best Local Similarity: 30.37% Mismatches: 53

Query Match: 15.14% Indels: 24

DB: 1 Gaps: 3

US-08-957-709A-71 (1-156) x AI395397 (1-491)

Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40

Db 449 TCTCTCTTTCGCGCTGGTTCAGATCTTTACGCTGCTCATGAAGCT----- 405

Qy 41 LeuIleAspValGluLysGluGlyLysValValleProProArgGluTyrAlaLeuIle 60

Db 404 -----ACCATCTCTCTGGGGGTCGTGAATTAATT 375

Qy 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80

Db 374 AAAACTGACATACAGATTGAATTCGCCGAAGGCTGTTATGGTCGCGTCGCTCCAGAGC 315

Qy 81 SerLeuAla---ArgGluGlyValIleGlySerPheAlaTrpPheAlaTrpAspProGlyTyrAsp 99

```

Db      314  GGACTAGCCTTAAACAGGGGATTGATGTGGTGGTGGTGTATTATACCGCGACTATAGG 255
Qy      100  GlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGly 119
Db      254  GGCAATGTTGGTGTAGTACTTTTCAACTTTGGAGAACCCAGAAATTTGAAATCAGAAAGCGT 195
Qy      120  GluArgPheValGlnIleAlaPheIleArgLeu-----Glu 131
Db      194  GACCGGTGACCAACTGATTGTGACGATATTTTCTGCCTGAACGTGATGGAATGTGAA 135
Qy      132  GlyProAlaArgAsnProTyrArgGlyAsnTyrGlnGlySerThr 146
Db      134  TCGTTCGGGAGACCGATCGTGTTCAAATGGTTATGGCTCAACT 90

RESULT 7
CD145288/c
LOCUS
DEFINITION
MGI-0104U-A343-G06-G, mRNA sequence.
CD145288
CD145288.1 GI:34683105
EST.
Schistosoma mansoni
Schistosoma mansoni
Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeida; Schistosomatidae; Schistosomatidae; Schistosoma.
1 (bases 1 to 507)
Verjovski-Almeida.S., DeMarco.R., Martins.E.A.L., Guimaraes,P.E.M.,
Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjo@iq.usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL http://bioinfo.iq.usp.br/schisto/
Plate: MGI-0104U-A343 row: 6 column: G.
FEATURES
source
1..507
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MGI-0104U-A343-G06.G"
/sex="mixed pool"
/dev_stage="germball"
/lab_host="Biomphalaria glabrata"
/clone_lib="MGI-0104"
/note="Vector: pGEM T-easy"
ORIGIN
Alignment Scores:
Pred. No.: 4,21e-05 Length: 507
Score: 122.00 Matches: 36
Percent Similarity: 47.37% Conservative: 18
Best Local Similarity: 31.58% Mismatches: 44
Query Match: 15.14% Indels: 16

```

```

DB:
US-08-957-709a-71 (1-156) x CD145288 (1-507)
Gaps: 2

Qy      21  SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
Db      305  TTCTCTCTTGGCGCTGGTTACGATCTTTAGCTGCTCATGAAGCT-----261
Qy      41  LeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyrAlaLeuIle 60
Db      260  -----ACCATACCTCTCTGGGGGCTCGTGAATTAATT 231
Qy      61  LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80
Db      230  AAAACTGCATACATGATTAATTCGCGAAGCTGTTATGGTCGCGCTCCACGAAGC 171
Qy      81  SerLeuAla---ArgGluGlyValIleGlySerPheAlaTyrValAspProGlyTyrAsp 99
Db      170  GGGCTAGCCTTAAACAGGGGATGATGTTGGTGGTGGTGTATTGACCGCGACTATAGG 111
Qy      100  GlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGly 119
Db      110  GGCAATGTTGGTGTAGTACTTTTCAACTTTGCAGAACAGAAATTTGAAATCAGAAAGGT 51
Qy      120  GluArgPheValGlnIleAlaPheIleArgLeuGluGlyPro 133
Db      50  GACCGGTGACCAACTGATTTGTGAACGTATTTTCTGTCTCT 9

RESULT 8
CB335318/c
LOCUS
DEFINITION
Tc005E10R Tribolium castaneum embryonic cDNA library Tribolium
castaneum cDNA clone Tc005E10 5', mRNA sequence.
CB335318
CB335318.1 GI:40543043
EST.
Tribolium castaneum (red flour beetle)
Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
Tenebrionidae; Tribolium.
1 (bases 1 to 530)
Savard,J. and Tautz,D.
A Tribolium castaneum EST project
Unpublished (2003)
Contact: Savard, J.
Abteilung für Evolutionsgenetik, AG Tautz
Institut für Genetik, Universität zu Köln
Weyertal 121, 50931 Köln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Seq primer: ML3R.
FEATURES
source
1..530
/organism="Tribolium castaneum"
/mol_type="mRNA"
/strain="Wild type"
/db_xref="taxon:7070"
/clone="Tc005E10"
/dev_stage="Mixed embryonic stages"
/clone_lib="Tribolium castaneum embryonic cDNA library"
/note="Vector: pBluescript SK; Site 1: EcoRI; Site 2:
XhoI; Uni-ZAP XR cDNA library (Stratagene) constructed by
Reinhard Schroder (1995)"
ORIGIN
Alignment Scores:
Pred. No.: 4,48e-05 Length: 530
Score: 122.00 Matches: 43
Percent Similarity: 49.62% Conservative: 23
Best Local Similarity: 32.33% Mismatches: 44
Query Match: 15.14% Indels: 24

```

```
DB: 6 Gaps: 4
US-08-957-709a-71 (1-156) x CB335318 (1-530)
Qy 16 PropheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAla 35
Db 503 CCTCCGACCAAGGTTGCTTTAAAGCCGCGGTATGATCTG-----AAGAGCGCT 453
Qy 36 PheValLysGlyLysLeuLeuAspValGluLysGluGlyLysValValLeuValLeuProArg 55
Db 452 TTT-----GACTGGTGGTCCCGGCCGGA 429
Qy 56 GluTyrAlaLeuLeuLeuThrLeuGluArgGlyLeuLysLeuProAspAspValMetGlyAsp 75
Db 428 GGTAGGCCCTCGTGACACCGGCATCAATCCAGCTCCCGAAGGTTGCTACGGACGG 369
Qy 76 MetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaTrp---Val 94
Db 368 ATCGCCCTAGTTCGGGTCTAGCAGTCAAGAAATTTTCATCGATGTTGGCGCGGGTCTGTG 309
Qy 95 AspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProVal 114
Db 308 GATGAGGACTACAGGGGGTCTCAAGGTGGTCTTATTCAACCATTCAGATCTCCGTTT 249
Qy 115 GluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAla 134
Db 248 GNAGTCCAGAGTGGCGACCGCATTCACAGCTAATTTGTGAGCGGATC----- 201
Qy 135 ArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArg 147
Db 200 -----TATTA-CCCGATATCGAGGAGTTCAAGA 172

RESULT 9
CK315881 800 bp mRNA linear EST 01-MAR-2004
LOCUS SB02028A1F02.f1 normalized Keck-Tagu Library SB02 Taeniopygia
DEFINITION guttata cDNA clone SB02028A1F02.f1 5, mRNA sequence.
ACCESSION CK315881
VERSION CK315881.1 GI:44825455
KEYWORDS EST.
SOURCE Taeniopygia guttata
ORGANISM Taeniopygia guttata
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Passeriformes; Estrilidae;
Estrildinae; Taeniopygia.
1 (bases 1 to 800)
Clayton,D.F., Arnold,A.P., Ball,G.F., Brenowitz,E., George,J.M.,
Mello,C.V., Wade,J., Replogle,K., Lewin,H., Band,M., Hernandez,A.
and Liu,L.
The Songbird Neurogenomics Initiative: An Evolving Public Resource
for Study of Genes, Brain, and Behavior
Unpublished (2004)
JOURNAL COMMENT
Contact: David F. Clayton
University of Illinois
B107 CILSL, 601 S. Goodwin, Urbana, IL 61801, USA
Tel: 217 244 3668
Fax: 217 244 1648
Email: dclayton@uiuc.edu
Base Calling/Quality Scores: PHRED from Washington University
Genome Center.
Vector Trimming: Cross match from Washington University Genome
Center PHRAP suite. Low quality bases (Phred score < 20) were
trimmed from both ends of the sequence by an in-house script.
This sequence is vector free and at least 200 bp in length. Funded
by PHS grant # R01 NS045264, 'Songbird Neurogenomics Initiative.'
PCR Primers
FORWARD: TAATACGACTCACTATAGG(T7)
BACKWARD: ATTAACTCCCTCACTAAAG(T3)
Insert Length: 800 Std Error: 0.00
Plate: SB02028A1 row: F column: 02
Seq primer: TAATACGACTCACTATAGG (T7)
High quality sequence stop: 800.
Location/Qualifiers
FEATURES
```

```
source
1..800
/morganism="Taeniopygia guttata"
/mol_type="mRNA"
/db_xref="taxon:59729"
/clone="SB02028A1F02.f1"
/tissue_type="brain"
/dev_stage="late embryo, post-hatch days 1, 10, 20, 45,
and adult (pooled)"
/lab_host="DH10B"
/clone_lib="normalized Keck-Tagu Library SB02"
/note="Organ: brain; Vector: pBS II SK(+); Site_1:
EcoRI(5' side of insert); Site_2: NotI (3' side of
insert); The library was constructed and normalized as
described by Bonaldo, M.F., Lennon, G. and Soares, M.B.
(1996). Genome Research 6(9): 791-806. An identifying tag
was added at the 3' during cDNA synthesis:
insertAAAAAAAAAAAAAAAAATGCGA."

ORIGIN
Alignment Scores:
Pred No.: 0.000194 Length: 800
Score: 119.00 Matches: 37
Percent Similarity: 46.15% Conservative: 17
Best Local Similarity: 31.62% Mismatches: 47
Query Match: 14.76% Indels: 16
DB: 7 Gaps: 2

US-08-957-709a-71 (1-156) x CK315881 (1-800)
Qy 18 SerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal 37
Db 115 TCCCGGGGCTCCGCGGCTGCTGCGGGTACGATCTGACAGTGCCTATGACTGT----- 168
Qy 38 LysGlyLysLeuIleAspValGluLysGluGlyLysValIleProProArgGluTyr 57
Db 169 -----GTGATACCAACCCATCGAGGAAG 189
Qy 58 AlaLeuIleLeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLys 77
Db 190 GCTGTGGTGAACACAGACATTCATAGCGCTTCTCTCGATGCTATGGCCGAGTAGCA 249
Qy 78 IleArgSerLeuAlaArgGluGlyValIleGlySerPheAlaTrpVal---AspPro 96
Db 250 CCACGTTCTGTTAGCTGCAAGACACTTCATAGACGTTGAGCTGGTGTATTGATGAG 309
Qy 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db 310 GATTACAGGGGAAATGTTGGTGTGCTGTTCACTTTGGCAAGGAGACATTTGAAGTT 369
Qy 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyPro 133
Db 370 AAAAAAGAGATAGAAATGCCAGCTCATCTGTGACGCACTTTATTCCT 420

RESULT 10
CK311070 826 bp mRNA linear EST 01-MAR-2004
LOCUS SB02007B2G06.f1 normalized Keck-Tagu Library SB02 Taeniopygia
DEFINITION guttata cDNA clone SB02007B2G06.f1 5, mRNA sequence.
ACCESSION CK311070
VERSION CK311070.1 GI:44820644
KEYWORDS EST.
SOURCE Taeniopygia guttata
ORGANISM Taeniopygia guttata
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Passeriformes; Estrildidae;
Estrildinae; Taeniopygia.
1 (bases 1 to 826)
Clayton,D.F., Arnold,A.P., Ball,G.F., Brenowitz,E., George,J.M.,
Mello,C.V., Wade,J., Replogle,K., Lewin,H., Band,M., Hernandez,A.
and Liu,L.
The Songbird Neurogenomics Initiative: An Evolving Public Resource
for Study of Genes, Brain, and Behavior
Unpublished (2004)
JOURNAL COMMENT
```


COMMENT

Contact: David F. Clayton
University of Illinois
Bio7 CUSL, 601 S. Goodwin, Urbana, IL 61801, USA
Tel: 217 244 3668
Fax: 217 244 1648
Email: dclayton@uiuc.edu
Base Calling/Quality Scores: PHRED from Washington University
Genome Center.

Vector Trimming: Cross_match from Washington University Genome
Center PHRAP suite. Low quality bases (Phred score < 20) were
trimmed from both ends of the sequence by an in-house script.
This sequence is vector free and at least 200 bp in length. Funded
by PHS grant # R01 NS045264, 'Songbird Neurogenomics Initiative.'

PCR Primers

FORWARD: TAATACGACTCACTATAGG (T7)

BACKWARD: ATTACCCCTCACTAAAG (T3)

Insert Length: 826 Std Error: 0.00

Plate: SB02007B2 row: G column: 06

Seq primer: TAATACGACTCACTATAGG (T7)

High quality sequence stop: 826.

FEATURES

source

1. .826

Location/Qualifiers

/organism="Taeniopygia guttata"

/mol_type="mRNA"

/db_xref="taxon:59729"

/clone="SB02007B2G06.f1"

/tissue_type="brain"

/dev_stage="late embryo, post-hatch days 1, 10, 20, 45,

and adult (pooled)"

/lab_host="DH10B"

/clone_lib="normalized Keck-Tagu Library SB02"

/note="Organ: brain; Vector: pBS II SK(+); Site: 1;

ECORI(5', side of insert); Site 2: NotI (3', side of

insert); The library was constructed and normalized as

described by Bonaldo, M.F., Lennon, G. and Soares, M.B.

(1996). Genome Research 6(9): 791-806. An identifying tag

was added at the 3' during cDNA synthesis:

InsertAAAAAAAAAAAAAAAAAATGCGCA."

ORIGIN

Alignment Scores:
Pred. No.: 0.000203 Length: 826
Score: 119.00 Matches: 37
Percent Similarity: 46.15% Conservative: 17
Best Local Similarity: 31.62% Mismatches: 46
Query Match: 14.76% Indels: 16
DB: 7 Gaps: 2

US-08-957-709a-71 (1-156) x CK311070 (1-826)

QY 18 SerGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal 37
Db 172 TCCCGGGGCTCCCGCGTCTCGGGCTACGATCTGTACAGTGCCTATGACTGT----- 225
QY 38 LysGlyLysLeuLeuAspValGluLysGluGlyLysValValleProProArgGluTyr 57
Db 226 -----GTGATACCAACCCATGGAGAG 246
QY 58 AlaLeuLeuLeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLys 77
Db 247 GCTGTGGTGAACACAGACATTCAATAGACATCTCTTCTGGATGCTATGCCGAGTAGCA 306
QY 78 IleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaTrpVal---AspPro 96
Db 307 CCACGTTCTGTGTAGCTCCAAAGCATTTTCATAGATGTTGGAGCTGGTGTATTGATGAG 366
QY 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db 367 GATTACAGGGAATGTTGGTGTGTACTGTTCACTTTGGCAAGGAGACGTTTGAAGTT 426
QY 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGlyPro 133
Db 427 AAAAAAGGAGTAGAATATGCCCGCTCATCTGTGAACGCATTATTATTCCT 477

RESULT 11

AA968575/c

LOCUS

DEFINITION

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

AA968575

```
Db 317 CCTCTGGGTGTTATGAAGAGTGCTCCACGGTCAGCGTTGGCTGCAAAACACTTTATT 258
Qy 89 GlySerPheAlaTrpVal---AspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyr 107
Db 257 GATGTAGAGCTGGTGCATAGATGAAGATTATAGAGGAATGTTGGTGTGTACTGTTT 198
Qy 108 AsnAlaSerAsnGluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPhe 127
Db 197 AATTTTGGCAAGAAAGATTGTAAGTCAAAAAGGTGATCGCAATTGCGACAGCTCATTTGC 138
Qy 128 IleArgLeuGluGlyPro 133
Db 137 GAACGGATTTTATCCA 120

RESULT 12
BZ549391/c
LOCUS BZ549391 839 bp DNA linear GSS 17-DEC-2002
DEFINITION pacsl-60_2014.xl pacsl-60 Pseudomonas aeruginosa genomic clone
ACCESSION pacsl-60_2014, genomic survey sequence.
VERSION BZ549391
KEYWORDS BZ549391.1 GI:27152972
SOURCE GSS.
ORGANISM Pseudomonas aeruginosa
REFERENCE Pseudomonas aeruginosa
AUTHORS Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
1 (bases 1 to 839)
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
J. Bacteriol. (2002) In press
JOURNAL Pseudomonas aeruginosa library
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
    Location/Qualifiers
        1..839
            /organism="Pseudomonas aeruginosa"
            /mol_type="genomic DNA"
            /strain="1-60"
            /db_xref="taxon:287"
            /clone="pacsl-60_2014"
            /clone_lib="pacsl-60"
            /note="clinical isolate 1-60 Whole genomic shotgun
            library."
ORIGIN
Alignment Scores:
Pred. No.: 0.00024 Length: 839
Score: 118.50 Matches: 31
Percent Similarity: 45.76% Conservative: 23
Best Local Similarity: 26.27% Mismatches: 35
Query Match: 14.70% Indels: 29
DB: 8 Gaps: 3

US-08-957-709A-71 (1-156) x BZ549391 (1-839)
Qy 13 LeuIleGluProPheSerGluGluSerLeuGlnProAla----- 25
Db 394 ATGATCGAGCCGTTCTGTCAGCGCCAGGTGCGCGCGGACGACGCGGGTGATTCC 335
Qy 26 -----GlyTyrAspLeuArgValGlyArgGlu----- 34
Db 334 TACGGGGTGTCCAGCTACGGCTACGACGTGCGCTCGCGCGCGGATTCAGGTGTTACC 275
Qy 35 -----AlaPheValLysGlyLysLeuIleAspVal 44
Db 274 AACATCCATTCCGGCGGTGTCGATCCGAAGAACTTCGACGAGAAAGCTTCGTCGACATC 215
```

```
Qy 45 GluLysGluGlyLysValValIleProArgGluTyrAlaLeuIleLeuThrLeuGlu 64
Db 214 AACAGCCGAC---GTCGTCATCATCCGCGCAACTCTTCGCTGGCGCGCACCGTCGAG 158
Qy 65 ArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSerSerLeuAlaArg 84
Db 157 TACTTCGCGATCCCGCGCGAGCTCTGCACCATCTGCTGGCGCAAGAGCACCTACGCGGT 98
Qy 85 GluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGlyAsnLeu 102
Db 97 TCGGCGATCATCTCAACGTCACCCGCTGGAGCGGAGTGGGAAGGCCAATGTA 44

RESULT 13
AJ451538
LOCUS AJ451538 663 bp mRNA linear EST 22-APR-2002
DEFINITION AJ451538 riken1 Gallus gallus cDNA clone 29b24r1, mRNA sequence.
ACCESSION AJ451538
VERSION AJ451538.1 GI:20261634
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
TITLE Phasianinae; Gallus.
JOURNAL Buerstedde, J.M.
COMMENT Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institut
Martinstr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
    Location/Qualifiers
        1..663
            /organism="Gallus gallus"
            /mol_type="mRNA"
            /db_xref="taxon:9031"
            /clone="29b24r1"
            /cell_type="bursal lymphocyte"
            /dev_stage="2-3 weeks old"
            /clone_lib="riken1"
            /note="CB inbred strain"
ORIGIN
Alignment Scores:
Pred. No.: 0.00027 Length: 663
Score: 117.00 Matches: 37
Percent Similarity: 47.01% Conservative: 18
Best Local Similarity: 31.62% Mismatches: 46
Query Match: 14.52% Indels: 16
DB: 1 Gaps: 2

US-08-957-709A-71 (1-156) x AJ451538 (1-663)
Qy 18 SerGluGluSerLeuLeuProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal 37
Db 136 TCCAAGGCTCTGCACGGCGCGGCTACACTTGTACACTGCTATGAC----- 186
Qy 38 LysGlyLysLeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyr 57
Db 187 -----TACGTGATCCCAACCATGGAACAG 210
Qy 58 AlaLeuIleLeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLys 77
Db 211 GCTGTAGTGAACACAGATTTCAGATTGCATCTCTCTGCTGGATGCTATGGCGGATGACA 270
Qy 78 IleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaTrpVal---AspPro 96
Db 271 CCGGCTTCTGTTAGTTCGTCGCAAGCACCTTCATAGATGTTGGTGTGCTGTTATTATGATGAG 330
Qy 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
```

```
Db 331 GATTACAGGGAATGTTGGCGTGGTACTTCAACTTGGCAGAGACTTTTGAAGTT 390
Qy 117 ArgTyTGlyGluArgPheValGlnLeuAlaPheIleArgLeuGluGlyPro 133
Db 391 AAGAAAGGAGATAGATTGCCCGCTCATCTGTGAACGCAATTTCTATCCT 441

RESULT 14
CO297884 481 bp mRNA linear EST 25-JUN-2004
LOCUS EK170827.5prme Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
DEFINITION melanogaster cDNA clone EK170827 5, mRNA sequence.
ACCESSION CO297884
VERSION 1
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 481)
AUTHORS Koczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
Peterson,E. and Swimmer,C.
TITLE Exelixis FlyTag EST Project CK01 Library
JOURNAL Unpublished (2004)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: EK1708 row: C column: 3
High quality sequence stop: 475.
FEATURES
Location/Qualifiers
source 1..481
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="EK170827"
/clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
/note="Organ: mixed stage embryos, imaginal disks, and
adult heads; Vector: pCDNA-SK+; Site:1: NotI; Site:2:
XhoI; Random primed, normalized library from mixed stage
embryos, imaginal disks, and adult heads."
ORIGIN
Alignment Scores:
Pred. No.: 0.0002 Length: 481
Score: 116.50 Matches: 35
Percent Similarity: 48.76% Conservative: 24
Best Local Similarity: 28.93% Mismatches: 39
Query Match: 14.45% Indels: 23
DB: 7 Gaps: 3

US-08-957-709A-71 (1-156) x CO297884 (1-481)
Qy 18 SerGluGluSerLeuGlnPro-----AlaGlyTyrAspLeuArg 30
Db 156 ACCGAGATGCTTTGGCGGTGGAGGGATCCGCCAAGCGCGGAGTTGACCTGCGC 215
Qy 31 ValGlyArgGluAlaPheValLysGlyLysLeuLeuAspValGluGlyLysVal 50
Db 216 AGCGCTACGAC-----GTT 230
Qy 51 ValIleProProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAsp 70
Db 231 GTGGTCCCGCAGCGCGGAAGGCCATTGTCAAGACCGCATCTGCAAGTCGAGGTTCCGGAG 290
Qy 71 AspValMetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySer 90
Db 291 GGCTCTCTACGACGCGCGCCACCATCCGGCTGGCGATCAAGAACTTCATTGATG 350
Qy 91 PheAlaTyr---ValAspProGlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAla 109
```

```
Db 351 GCGCCGGTGTGTGGACGAGATTATCGGCAATCTCGGCTGCTCTGTTCAATCAC 410
Qy 110 SerAsnGluProValGlnLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArg 129
Db 411 TCAGATGTTGATTTCGAGGTGAAGCATGCGACCGCATCGCCAGTTCATTGTCGAGCCT 470
Qy 130 Leu 130
Db 471 ATC 473

RESULT 15
CO297326 444 bp mRNA linear EST 01-OCT-2004
LOCUS 3GAL 20E01 57, mRNA sequence.
DEFINITION CV037326
ACCESSION CV037326
VERSION 1
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 444)
AUTHORS Evock-Clover,C.M., Ashwell,C.M., McMurtry,J.P., Lillehoj,H.S.,
Matukumalli,L.K. and Van Tassel,C.P.
TITLE Characterization of expressed sequence tags generated from multiple
chicken tissues
JOURNAL Unpublished (2004)
COMMENT Contact: Christina M. Clover
Growth Biology Laboratory
Animal and Natural Resources Institute
Bldg. 200 Rm215 BARC-East, Beltsville, MD 20705, USA
Tel: 3015048224
Fax: 3015048623
Email: chris@anri.barc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim_alt -, -trim_fasta. Vector identified
by cross match using options -minmatch 12 -minscore 18
Plate: 20 row: E column: 01
Seq primer: CCAGTCACGAGCTGTGTAACGC
High quality sequence stop: 444.
FEATURES
Location/Qualifiers
source 1..444
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Leghorn and broiler"
/db_xref="taxon:9031"
/clone="3GAL 20E01"
/lab_host="DH5alpha"
/clone_lib="BARC 3GAL chicken mixed tissue"
/note="Vector: pBluescript, SK+, Stratagene; Site:1: NotI;
Site:2: EcoRI; Normalized library of pooled RNA isolated
from whole brain, ultimobranchial gland, parathyroid
gland, cecal tonsil and primordial germ cells Multiple"
ORIGIN
Alignment Scores:
Pred. No.: 0.000208 Length: 444
Score: 116.00 Matches: 37
Percent Similarity: 47.01% Conservative: 18
Best Local Similarity: 31.62% Mismatches: 46
Query Match: 14.33% Indels: 16
DB: 7 Gaps: 2

US-08-957-709A-71 (1-156) x CV037326 (1-444)
Qy 18 SerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal 37
Db 136 TCCAAAGGCTCTGCGACGGCGCGGTACACTGTACAGTGCCTATGAC----- 186
Qy 38 LysGlyLysLeuIleAspValGluLysGluGlyLysValIleProProArgGluTyr 57
```

```
Db 187 -----TACGTGATCCCAACCATGAAAAAG 210
Qy 58 AlaLeuIleLeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLys 77
Db 211 GCTGTAGTGAACACAGACATTGCACCTTCTGCTGGATGCTATGGCCGAGTAGCA 270
Qy 78 IleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaTrpVal---AspPro 96
Db 271 CCGGTTCTGCTTAGCTGCAAGCAGCACTTCATAGATGTTGGTGGTGTATTATTGATGAG 330
Qy 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db 331 GATTACAGGGGAAATGTTGGCGTGGTACTCTTCAACTTTGGCAAGGAGACCTTTGAAGTT 390
Qy 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyPro 133
Db 391 AAGAAAGGAGATAGAAATTGGCCAGCTCATCTGTGAACGCATTTTCTATCCT 441
```

Search completed: July 2, 2005, 11:06:29
Job time : 2827 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 2, 2005, 10:23:22 ; Search time 524 Seconds
(without alignments)
1867.291 Million cell updates/sec

Title: US-08-957-709a-71
Perfect score: 806
Sequence: 1 MLLPWKIRKEILIEPFSEE.....PYRGNVQGSRLAFSRKKKL 156

Scoring table:
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-DB=Published Applications NA -QWTF=faGap -SUPP=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MAIRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USRR=US08957709 @CGN 1.1 740 @runat 01072005 154633_26120
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCk=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	214	26.6	567	9	US-09-738-626-3140	Sequence 3140, Ap
2	214	26.6	3309400	9	US-09-738-626-1	Sequence 1, Appli
3	210	26.1	573	15	US-10-156-761-4448	Sequence 4448, Ap
4	210	26.1	9025608	15	US-10-156-761-1	Sequence 1, Appli
5	194	24.1	423	9	US-09-974-300-5011	Sequence 5011, Ap
6	191.5	23.8	2256646	19	US-10-470-565-1	Sequence 1, Appli
7	174	21.6	640681	9	US-09-790-988-1	Sequence 1, Appli
8	168.5	20.9	89047	18	US-10-672-787-34	Sequence 34, Appli
9	164.5	20.4	2731748	19	US-10-297-465A-1	Sequence 1, Appli
10	160.5	19.9	2964	20	US-10-784-986-3	Sequence 3, Appli
11	154	19.1	1230025	17	US-10-289-762-1	Sequence 1, Appli
12	149.5	18.5	586	19	US-10-437-963-14349	Sequence 14349, A
13	149.5	18.5	1830121	17	US-10-329-670-1	Sequence 1, Appli
14	149.5	18.5	1830121	20	US-10-158-865-1	Sequence 1, Appli
15	149.5	18.5	1830121	22	US-10-981-687-1	Sequence 1, Appli
16	145.5	18.1	3129	17	US-10-282-122A-13918	Sequence 13918, A
17	134.5	16.7	534	10	US-09-975-719-246	Sequence 246, App
18	134.5	16.7	42235	10	US-09-975-719-1	Sequence 1, Appli
19	132.5	16.4	501	10	US-09-975-719-244	Sequence 244, App
20	125.5	15.6	4658	21	US-10-258-089-16	Sequence 16, Appli
21	125.5	15.6	4658	21	US-10-258-089-27	Sequence 27, Appli
22	125.5	15.6	10112	14	US-10-239-804-12	Sequence 12, Appli
23	125.5	15.6	10112	14	US-10-239-804-70	Sequence 70, Appli
24	125.5	15.6	10112	15	US-10-134-643-6	Sequence 6, Appli
25	125.5	15.6	10112	20	US-10-838-906-24	Sequence 24, Appli
26	125.5	15.6	10112	21	US-10-258-089-5	Sequence 5, Appli
27	125.5	15.6	10114	14	US-10-239-804-13	Sequence 13, Appli
28	125.5	15.6	10114	14	US-10-239-804-71	Sequence 71, Appli
29	125.5	15.6	10114	20	US-10-838-906-25	Sequence 25, Appli
30	125.5	15.6	10114	21	US-10-258-089-18	Sequence 18, Appli
31	125.5	15.6	10227	21	US-10-258-089-6	Sequence 6, Appli
32	125.5	15.6	10392	21	US-10-258-089-17	Sequence 17, Appli
33	125.5	15.6	10815	15	US-10-134-643-10	Sequence 10, Appli
34	125.5	15.6	10815	21	US-10-258-089-7	Sequence 7, Appli
35	125.5	15.6	10930	21	US-10-258-089-8	Sequence 8, Appli
36	125.5	15.6	11874	15	US-10-134-643-1	Sequence 1, Appli
37	125.5	15.6	12481	14	US-10-239-804-5	Sequence 5, Appli
38	125.5	15.6	12481	14	US-10-239-804-69	Sequence 69, Appli
39	125.5	15.6	12481	15	US-10-134-643-9	Sequence 9, Appli
40	125.5	15.6	12481	20	US-10-838-906-22	Sequence 22, Appli
41	125.5	15.6	12481	21	US-10-258-089-12	Sequence 12, Appli
42	121	15.0	441	21	US-10-472-928-4751	Sequence 4751, Ap
43	121	15.0	477	22	US-10-617-320-924	Sequence 924, App
44	121	15.0	2162598	21	US-10-472-928-4979	Sequence 4979, Ap
45	117	14.5	10223	8	US-08-961-527-73	Sequence 73, Appli

ALIGNMENTS

RESULT 1
US-09-738-626-3140
; Sequence 3140, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOHI, AKIHIRO
; APPLICANT: IKEDA, NASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

```

, FILE REFERENCE: 249-125
, CURRENT APPLICATION NUMBER: US/09/738,626
, CURRENT FILING DATE: 2000-12-18
, PRIOR APPLICATION NUMBER: JP 99/377484
, PRIOR FILING DATE: 1999-12-16
, PRIOR APPLICATION NUMBER: JP 00/159162
, PRIOR FILING DATE: 2000-04-07
, PRIOR APPLICATION NUMBER: JP 00/280988
, PRIOR FILING DATE: 2000-08-03
, NUMBER OF SEQ ID NOS: 7059
, SOFTWARE: PatentIn ver. 3.0
, SEQ ID NO 3140
, LENGTH: 567
, TYPE: DNA
, ORGANISM: Corynebacterium glutamicum
US-09-738-626-3140

```

Alignment Scores:	8.32e-20	Length:	567
Pred. No.:	Score:	Matches:	56
	214.00	Conservative:	29
Percent Similarity:	50.30%	Mismatches:	60
Best Local Similarity:	33.14%	Indels:	24
Query Match:	26.55%	Gaps:	4
DB:	9		

US-08-957-709A-71 (1-156) x US-09-738-626-3140 (1-567)

Qy	1	MetLeuLeuProAspTrpIysIleArgLys-----GluIleLeuIleGlu 15
		:::
Db	1	GTGCTCTCTTTTCAGATCGTGACATTCGTAATAATCAATGCACGCGCGCTTTGGGAATGGAA 60
Qy	16	ProPheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
Db	61	CGTTTCGACGCTGAGCTGATCTCAGCGCGTCGAGTGTCCGATGGACCGCTACTTTC 120
Qy	34	GluAlaPheValLysGlyLys----- 40
		:::
Db	121	CGGGTTTTCATAAATACTCTAAAGTACACCCACACATTGACCCCTAAGTTGAATCAGGATGAGCTG 180
Qy	41	-----LeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyrAla 58
Db	181	ACCAGCCTTGTGAGGTTGAGACCGCGGAGGCTTTGTGCTGCATCCGGGTGAGTTGTG 240
Qy	59	LeuIleLeuThrLeuGluArgIleLysLysLeuProAspValMetGlyAspMetLysIle 78
Db	241	CTGGCGTCCACGCTGGAAAGTTCACTTTGGCTGCGCATCTGGCTGGTCTGTTGGAGGGT 300
Qy	79	ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTyrValAspPro 96
Db	301	AGTGTGCTCTTTGGTGGCTTGGCTTGTGTGACGCACTCTACTGTGTGGTTTCATTGATCCT 360
Qy	97	GlyTyrAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db	361	GGTTTTAGTGGTTACATCACGTTGAGATTGCCAATGTGGCTAATCTGCCGAGTACACGTTG 420
Qy	117	ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaAlaArgAsn 136
Db	421	TGCCCGGGTATGAAGTGGGGCAGTTGGCTTTGTTCCAGATGAGTTCCTCTGGCGAGACT 480
Qy	137	ProTyrArgGlyAsnTyrGlnGlySer 145
		:::
Db	481	CGGTATGGTTCCGGCAAGCTTCGTTTCG 507

RESULT 2

US-09-738-626-1/C
; Sequence i, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OKIAI, KEIICHI

```

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4448
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(573)
US-10-156-761-4448

Alignment Scores:
Pred. No.: 3,06e-19 Length: 573
Score: 210.00 Matches: 53
Percent Similarity: 50.30% Conservative: 32
Best Local Similarity: 31.36% Mismatches: 60
Query Match: 26.05% Indels: 24
DB: 15 Gaps: 3

US-08-957-709a-71 (1-156) x US-10-156-761-4448 (1-573)

QY 1 MetLeuLeuProAspTrpLysIleArgLysGluLe-----LeuIleGlu 15
DB 1 GTGCTTCTCTCAGCAAGACATCCGGCGGCGAGATCGACGCCGAGCGGTTCGATATCGAC 60
QY 16 PropSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
DB 61 CCGTACGAGATCCATTCGTGGTGCAGCCCTCGAGCATCGATGTGCGGTTCGACCGCTACTTC 120
QY 34 -----GluAlaPheValLys 38
DB 121 CGGGTGTTCGAGAAATCACCGCTATCCGCATATCGACCCCTTCCTCGAGCAGCGCGATCTG 180
QY 39 GlyLysLeuLeuAspValGluLysGluGlyLysValValIleProProArgGluTyrAla 58
DB 181 ACGCGTCTCGTCCGAGCCGAGGCGAGCGAGCCGCTTCATCTTCATCCCGGGGAGTTTCGTC 240
QY 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIle 78
DB 241 CTCGCCAGTACCTACGAGGTCACTCTCCCTCCCGACGACCTCGCGTCCGGCTGGAGGGG 300
QY 79 ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
DB 301 AAGAGTTTCCTCGGGCGGTCTGGGCTGGTGGTCAACCATCTCCACCGCCGGGTTCATCGATCC 360
QY 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
DB 361 GGGTTCTCCGGGCACGTCACGCTCGAGCTCTCCAACTCCGCCACCCCTTCGATCAAGCTC 420
QY 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
DB 421 TGGCCGGGATGAAGATCGGGCAGCTGTGCATGTTCCGGCTCAGCTCGCCCGCGGATTC 480
QY 137 ProTyrArgGlyAsnTyrGlnGlySer 145
DB 481 CCCTACGGCAGTACCGCTACGGTTTC 507

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 1.86e-13 Length: 9025608
Score: 210.00 Matches: 53
Percent Similarity: 50.30% Conservative: 32
Best Local Similarity: 31.36% Mismatches: 60
Query Match: 26.05% Indels: 24
DB: 15 Gaps: 3

US-08-957-709a-71 (1-156) x US-10-156-761-1 (1-9025608)

QY 1 MetLeuLeuProAspTrpLysIleArgLysGluLe-----LeuIleGlu 15
DB 5453355 GTGCTTCTCTCAGCAAGACATCCGGCGGCGAGATCGACGCCGAGCGGTTCGATATCGAC 5453414
QY 16 PropSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
DB 5453415 CCGTACGAGATCCATTCGTGGTGCAGCCCTCGAGCATCGATGTGCGGTTCGACCGCTACTTC 5453474
QY 34 -----GluAlaPheValLys 38
DB 5453475 CGGGTGTTCGAGAAATCACCGCTATCCGCATATCGACCCCTTCCTCGAGCAGCGCGATCTG 5453534
QY 39 GlyLysLeuLeuAspValGluLysGluGlyLysValValIleProProArgGluTyrAla 58
DB 5453535 ACGCGTCTCGTCCGAGCCGAGGCGAGCGCGCTTCATCTTCATCCCGGGGAGTTTCGTC 5453594
QY 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIle 78
DB 5453595 CTCGCCAGTACCTACGAGGTCACTCTCCCTCCCGACGACCTCGCGTCCGGCTGGAGGGG 5453654
QY 79 ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
DB 5453655 AAGAGTTTCCTCGGGCGGTCTGGGCTGGTGGTCAACCATCTCCACCGCCGGGTTCATCGATCC 5453714
QY 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
DB 5453715 GGGTTCTCCGGGCACGTCACGCTCGAGCTCTCAACCTCGGCCACCCCTTCGATCAAGCTC 5453774
QY 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
DB 5453775 TGGCCGGGATGAAGATCGGGCAGCTGTGCATGTTCCGGCTCAGCTCGCCCGCGGATTC 5453834
QY 137 ProTyrArgGlyAsnTyrGlnGlySer 145
```



```

; PRIOR FILING DATE: 2001-06-17
;
; NUMBER OF SEQ ID NOS: 1
;
; SOFTWARE: PatentIn version 3.2
;

```

; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465A-1

Alignment Scores:

Pred. No.: 8.38e-08 Length: 2731748
Score: 164.50 Matches: 43
Percent Similarity: 48.23% Conservative: 25
Best Local Similarity: 30.50% Mismatches: 52
Query Match: 20.41% Indels: 21
DB: 19 Gaps: 4

US-08-957-709A-71 (1-156) x US-10-297-465A-1 (1-2731748)

QY 26 GlyTyrAspLeuArgValGlyArgGlu----- 34
DB 716579 GGCACGAGCGTGGCTGTTCCCGTGAATTCAGATATTACCAATATCAACTCAACCAATT 716520
QY 35 -----AlapheVallysglylylsLeulleAspValGluLysGluLysVal 50
DB 716519 GTCGATCCAAAACAATTCGATTAATGAAGCTTCATTGATGTCGAATCAGAC---GTCGTC 716463
QY 51 ValIleProProArgGluTyrAlaLeulleLeuThrLeuGluArgIleLysLeuProAsp 70
DB 716462 ATCATCCCAACCAACAGCTTCGCTTGGCAGCCACCATGTGATTTCGGCATCCACGC 716403
QY 71 AspValMetGlyAspMetLysIleArgSerSerLeuAlaArgGluValIleGlySer 90
DB 716402 AATGTACTGGTGATTTGTTCCGAAAGACACCTACGACGTTGTGAATCATCGTCAAT 716343
QY 91 PheAlaTrpValAspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSer 110
DB 716342 GTGACCCCAACGAGTGGGAAGGACACGTCACCTTGAAGTCAGCAACACCAACC 716283
QY 111 AsnGluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeu 130
DB 716282 CCGTTACCCGACGAGCATCTACCGAAGAGGGGTGTAGCCAGATGTTGTTCTCGAAGCC 716223
QY 131 -----GluGlyProAlaArgAsnProTyrArg-----GlyAsnTyrGlnGlySer 145
DB 716222 GATCCGACGAGCGTGTGCCAAACATCTACCGCGACCGCAACGCGCAAAATATCAAGGACAG 716163
QY 146 Thr 146
DB 716162 ACG 716160

RESULT 10

US-10-784-986-3
; Sequence 3, Application US/10784986
; Publication NO. US20040229311A1
; GENERAL INFORMATION:
; APPLICANT: HIRANO, Seiko
; APPLICANT: YASUEDA, Hisashi
; TITLE OF INVENTION: Novel lysine decarboxylase gene and method for
; FILE REFERENCE: producing L-lysine
; CURRENT APPLICATION NUMBER: US/10/784,986
; PRIOR APPLICATION NUMBER: 2004-02-25
; PRIOR FILING DATE: 2003-02-25
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; TYPE: DNA
; ORGANISM: Methylophilus methylotrophus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (684)..(2930)
US-10-784-986-3

Alignment Scores:

Pred. No.: 2.5e-11 Length: 2964
Score: 160.50 Matches: 44
Percent Similarity: 42.11% Conservative: 28
Best Local Similarity: 25.73% Mismatches: 68
Query Match: 19.91% Indels: 31
DB: 20 Gaps: 3

US-08-957-709A-71 (1-156) x US-10-784-986-3 (1-2964)

QY 13 LeulleGluProPheSerGluGluSerLeuGlnProAla----- 25
DB 61 ATGATTGAGCGGTTTGAGCCCAAGCTTGTACGTGAGACCAATGGCCAGAAGATTGTTCT 120
QY 26 -----GlyTyrAspLeuArg----- 30
DB 121 TATGGCACTCTTCTTACGGTTACGATATCGTTGTGTGACGAATTCGCGGTATTATTACC 180
QY 31 -----ValGlyArgGluAlaPheValLysGlyLysLeulleAspVal 44
DB 181 AATATCAACAGCACCATAGTTGACCCCAAGCAATTTGACCCGACGTCGTTTGTGAGGTC 240
QY 45 GluLysGluGlyLysValIleProProArgGluTyrAlaLeulleLeuThrLeuGlu 64
DB 241 TCCGCAAAAGGCTATTGCGTGATTCCCTTAATCATTTGCACTGCGCGCAGGTAGAG 300
QY 65 ArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSerSerLeuAlaArg 84
DB 301 TATTTCCGATTTCTTCGCTGTACTGTATGCTTCGCGCAAGTCGACTTATGCGCGT 360
QY 85 GluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGlyAsnLeuThrLeu 104
DB 361 TCGCGCATTAATCGTCAACGTCACCCCTTTGAACCAAGAGTGGAAAGGCTATGTCACACTA 420
QY 105 MetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGluArgPheValGln 124
DB 421 GAGTTTCAGCAACACCAACCGCTACCCGCAAAATTTATGCTGGCAAGGCTGTGCGCAA 480
QY 125 IleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArg-----GlyAsn 141
DB 481 GTGCTGTTCTTTGAGTCTGTGATAAATCTGTGAACGAGCTACAAAGACCGTGTGTTGTA 540
QY 142 TyrGlnGlySerThrArgLeuAlaPheSerLys 152
DB 541 TACCAGGCTCAAAATTTGGCGTGACCTCGCAAAA 573

RESULT 11

US-10-289-762-1
; Sequence 1, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature


```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (585001)..(600000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600001)..(615000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (615001)..(630000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (630001)..(645000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (645001)..(660000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (660001)..(675000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (675001)..(690000)
; OTHER INFORMATION: n=a or c or g or t
; FEATURE:
```

```
Alignment Scores:
Pred. No.: 8,248-07 Length: 1230025
Score: 154.00 Matches: 46
Percent Similarity: 43.72% Conservative: 34
Best Local Similarity: 25.14% Mismatches: 63
Query Match: 19.11% Indels: 40
DB: 17 Gaps: 6
```

US-08-957-709A-71 (1-156) x US-10-289-762-1 (1-1230025)

```
QY 13 LeuileGluProPhe-----SerGluGluSer----- 21
DB 450349 ATGATCCATCCCTTTGTTAATGCGCAAGTGAACGTAATAGGACAGCGGAAAAAAGCTT 450408
QY 22 -----LeuGlnProAlaGlyTyrAspLeuArgValGlyArgGlu----- 34
DB 450409 ATAAGTTACGGCCTATCGAGTTATGTTACGACCTCCGCTATCTCGAGAATTCAAAGTG 450468
QY 35 -----AlaPheValLysGlyLysLeuile 42
DB 450469 TTCACCAATGCTATAACTCTGTTGTTGATCCAAAATGCTTTACTGAGATATCTTCATC 450528
QY 43 AspValGluLysGluGlyLysValValilleProProArgGluTyrAlaLeuileLeuThr 62
DB 450529 TCTATT---ACTGATGACGCTGTATTGTTCTCCAAATTCCTTTGCTCTAGCTCGTAGC 450585
QY 63 LeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSerSerLeu 82
DB 450586 GTTGAGTATTTCCGAATTCCTAGAAAATGCTTTAACAATGTGTATAGGAAGTCTACATAT 450645
QY 83 AlaArgGluGlyValilleGlySerPheAlaTrpValAspProGlyTyrAspGlyValAsnLeu 102
DB 450646 GCACGCTGTGGAATATCGTAATATGTCACACCTTTTGGACCTGGAATGGGAAGGCGATGTG 450705
QY 103 ThrLeuMetLysAsnAlaSerAsnGluProValGluLeuArgTyrGlyGluArgPhe 122
DB 450706 ACTATAGAAATTTCTTAACACTAGCCATTCGCCAGCGAAAATTTACGCTAAATGAAGGGATT 450765
QY 123 ValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyr-----Arg 139
DB 450766 GCCCAGGCTTATTCTTTTGGAGTCTAGTAGACCTCGAGGTTTCTTATGACAGACAGAAA 450825
QY 140 GlyAsnTyrGlnGlySerThrArgLeu-----AlaPheSerLysArg 153
```

```
Db 450826 GGAAAGTATCAAAAGCAACAAAGGCATCACCGTACCTTGTGTCTAAAGTTTCAGTAAGAAA 450885
QY 154 LysLysLeu 156
DB 450886 AAAAAAATCG 450894
RESULT 12
US-10-437-963-14349
; Sequence 14349, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 14349
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_20296C.1
US-10-437-963-14349
```

```
Alignment Scores:
Pred. No.: 9,29e-11 Length: 586
Score: 149.50 Matches: 42
Percent Similarity: 50.00% Conservative: 18
Best Local Similarity: 35.00% Mismatches: 56
Query Match: 18.55% Indels: 4
DB: 19 Gaps: 2
```

US-08-957-709A-71 (1-156) x US-10-437-963-14349 (1-586)

```
QY 22 LeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLysLeu 41
DB 160 TTAAGTGGCGCTGTGAACAAGTCAGCCACAACTTGAAATAATCATGATGATGAAT 219
QY 42 IleAspValGluLysGluGlyLysValValilleProProArgGluTyrAlaLeuileLeu 61
DB 220 ATTTAGCAGAGGTGAAGCT---TTTTCTTCATCCAGCGGAATTAGCCTTGGCCACT 276
QY 62 ThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSerSer 81
DB 277 ACCCTAGAGTCAGTGAATAATTACCGGATAATATTGTTGGTTGGCTAGATGAACGTTCTTCA 336
QY 82 LeuAlaArgGluGlyValilleGlySerPheAla-----TrpValAspPro-GlyTrpAs 99
DB 337 TTGGCTCCCTTAGGATTAATGGTCATGTTACCGCTCATCGCATTTGATCCAGGTTGGCA 396
QY 99 pGlyAsnLeuThrLeuMetLysAsnAlaSerAsnGluProValGluLeuArgTyrGln 119
DB 397 CGGCAAAATTTGATTAGAAATTCCTTAATGCTGGTAAACTTACCATTAGCCTTCCGCTTAA 456
QY 119 yGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyr 138
DB 457 TATGCGCATGTTGTCATTAAAGCTTTTGAAGTATTGATGAGCAGCGGCTTAAGCCTTAT 514
```

```
RESULT 13
US-10-329-670-1
; Sequence 1, Application US/10329670
; Publication No. US20040018503A1
; GENERAL INFORMATION:
```

APPLICANT: Fleischmann et al.	OTHER INFORMATION: n equals a, t, g or c
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag	FEATURE:
TITLE OF INVENTION: Thereof, and Uses Thereof	NAME/KEY: misc feature
FILE REFERENCE: PB186P1	LOCATION: (51334)..(51334)
CURRENT APPLICATION NUMBER: US/10/329,670	OTHER INFORMATION: n equals a, t, g or c
CURRENT FILING DATE: 2002-12-24	FEATURE:
PRIOR APPLICATION NUMBER: US 08/643,990	NAME/KEY: misc feature
PRIOR FILING DATE: 2000-08-23	LOCATION: (51602)..(51602)
PRIOR APPLICATION NUMBER: US 08/487,429	OTHER INFORMATION: n equals a, t, g or c
PRIOR FILING DATE: 1995-06-07	FEATURE:
PRIOR APPLICATION NUMBER: US 08/426,787	NAME/KEY: misc feature
PRIOR FILING DATE: 1995-04-21	LOCATION: (51786)..(51786)
NUMBER OF SEQ ID NOS: 1	OTHER INFORMATION: n equals a, t, g or c
SOFTWARE: PatentIn version 3.1	FEATURE:
SEQ ID NO 1	NAME/KEY: misc feature
LENGTH: 1830121	LOCATION: (51805)..(51805)
TYPE: DNA	OTHER INFORMATION: n equals a, t, g or c
ORGANISM: Haemophilus influenzae	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (4747)..(4747)	LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (9921)..(9921)	LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (10150)..(10150)	LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (29298)..(29298)	LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (36543)..(36543)	LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (36551)..(36551)	LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (36636)..(36636)	LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (40808)..(40810)	LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (44416)..(44416)	LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (44905)..(44905)	LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (44975)..(44975)	LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (45593)..(45593)	LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (45732)..(45732)	LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c
FEATURE:	FEATURE:
NAME/KEY: misc feature	NAME/KEY: misc feature
LOCATION: (47036)..(47036)	LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c	OTHER INFORMATION: n equals a, t, g or c

NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature

LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)


```

; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a,t,c, or g
; NAME/KEY: misc feature
; LOCATION: (140398)..(140398)

```

```

; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:

Alignment Scores:
Pred. No.:      6.07e-06      Length:      1830121
Score:          149.50      Matches:      44
Percent Similarity: 45.95%      Conservative: 24
Best Local Similarity: 29.73%      Mismatches: 53
Query Match:      18.55%      Indels:      27
DB:                22      Gaps:          5

US-08-957-709a-71 (1-156) x US-10-981-687-1 (1-1830121)

QY 18 SerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheVal 37
Db 146725 AATAATGACAAATCAACGGTCGACAAATGATGTTCTGGGGAATCTTTCCGTGTA 146784
QY 38 -----LysGlyLysLeuIleAspValGlu-----LysGlu----- 47
Db 146785 TTTCGTGAACACTCTGCACCTTTTATTGATCTAAGCGGCGCCGAAAGAGATATCGCC 146844
QY 48 -----GlyLysValValIlePro----- 53
Db 146845 CAGCTTGAATCAGTAATAGCGATGAAATATTATTCAGAGGGGGAAGCATTTTCTTA 146904
QY 54 ---ProArgGluTyrAlaLeuIleLeuThrLeuGluArgIleLysLeuProAspVal 72
Db 146905 CATCTGTGTACTTTAGCCCTTAGCAACACCGTTGAATCCGTGAAATTCGACCAATATT 146964
QY 73 MetGlyAspMetLysIleArgSerSerLeuAlaArgGluGlyValIleGlySerPheAla 92
Db 146965 ATCGGTGGTTAGATGGCGTTCTTTTAGCCCGCTTAGGCTTAATGGTACACGTACC 147024
QY 93 -----TrpValAspProGlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSer 110
Db 147025 GCACACCGTATTGATCCAGTTGGGAAGAAAAATCGTCTGGAATTTTATAATTCAGG 147084
QY 111 AsnGluProValGluLeuArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeu 130
Db 147085 AAATTACCATTAGCATTAGCCCAAAATGGAATTTGGTGCCTTAAGTTTGAAGTATTA 147144
QY 131 GluGlyProAlaArgAsnProTyr 138
Db 147145 AGTGGGGAASSAAAAACGACCATAC 147168

```

Search completed: July 2, 2005, 13:18:17
Job time : 3989 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 2, 2005, 08:11:01 ; Search time 473 Seconds
(without alignments)
1952.388 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 806
Sequence: 1 MLLPWKIRKEILIEPFSEE.....PYRGNVQSGSTRLAFSKRKL 156

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: : 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US08957709/runat_01072005_154631_25999/app_query.fasta_1.327
-DB=N Geneseq 16Dec04 -QFMT=fastap -SUPEX=ring -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORF=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US08957709 @CNC 1 1 644 @runat_01072005_154631_25999 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 16Dec04: *
1: Geneseqm1980s: *
2: Geneseqm1990s: *
3: Geneseqm2000s: *
4: Geneseqm2001as: *
5: Geneseqm2001bs: *
6: Geneseqm2002as: *
7: Geneseqm2002bs: *
8: Geneseqm2003as: *
9: Geneseqm2003bs: *
10: Geneseqm2003cs: *
11: Geneseqm2003ds: *
12: Geneseqm2004as: *
13: Geneseqm2004bs: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	806	100.0	471	2 AAV63860	AAV63860 Polymeras
2	703	87.2	265118	5 AAH41227	AAH41227 Pyrococcu
3	696	86.4	110000	12 ADN46845	ADN46845 12
c 4	696	86.4	110000	12 ADN47591_08	Continuation (13 o
c 5	696	86.4	110000	12 ADN46123_12	Continuation (9 of

6	696	86.4	110000	12 ADN47209_08	Continuation (9 of
c 7	696	86.4	110000	12 ADN46464_12	Continuation (13 o
c 8	696	86.4	110000	12 ADN47960_08	Continuation (9 of
c 9	242	30.0	3382	2 AAQ62300	AAQ62300 p2am-l3 p
10	227.5	28.2	597	12 ADK16664	ADK16664 Nanoarcha
11	227.5	28.2	110000	12 ADK16049_2	Continuation (3 of
c 12	220	27.3	110000	11 ADM27081_02	Continuation (3 of
c 13	214	26.6	567	5 AAH68105	AAH68105 C glutami
c 14	214	26.6	309400	5 AAH68534	AAH68534 C glutami
c 15	214	26.6	349980	5 AAH68533	AAH68533 C glutami
16	196	24.3	580	4 AAF72258	AAF72258 Corynebac
17	194	24.1	423	6 ABK77720	ABK77720 Bacillus
18	194	24.1	110000	4 AAI99682_03	Continuation (4 of
c 19	194	24.1	110000	4 AAI99683_03	Continuation (4 of
c 20	191.5	23.8	349980	6 ABQ81849	ABQ81849 Bifidobac
c 21	185.5	23.0	594	9 ADA29252	ADA29252 DNA encod
22	181	22.5	2006	2 AAV90555	AAV90555 Nucleotid
23	181	22.5	2588	2 AAV90871	AAV90871 Nucleotid
24	174	21.6	110000	6 ABA92787_1	Continuation (2 of
25	172	21.3	609	11 ABD02487	ABD02487 Pseudomon
26	170.5	21.2	582	10 ACF68843	ACF68843 Photorhab
c 27	170.5	21.2	110000	10 ACF67367_16	Continuation (17 o
c 28	170.5	21.2	249878	12 ACF65381	ACF65381 Photorhab
c 29	168.5	20.9	579	12 ADL04075	ADL04075 DNA encod
30	168.5	20.9	615	10 ADF00606	ADF00606 Bacterial
31	168.5	20.9	89047	4 AAF28547	AAF28547 Genomic f
32	164	20.3	654	11 ACH95333	ACH95333 Klebsiell
33	164	20.3	110000	11 ADM27081_15	Continuation (16 o
34	160.5	19.9	2964	13 ADR89139	ADR89139 Nucleotid
35	156	19.4	564	10 ABZ39371	ABZ39371 N. gonorr
36	155.5	19.3	419	6 ABX66923	ABX66923 Helicobac
37	155	19.2	78845	3 AAA81463	AAA81463 N. mening
c 38	155	19.2	110000	3 AAA81490_08	Continuation (9 of
c 39	155	19.2	349980	3 AAF21608	AAF21608 Neisseria
40	154.5	19.2	573	12 ADQ57545	ADQ57545 Actinobac
41	154	19.1	110000	2 AAV21209_10	Continuation (11 o
42	154	19.1	110000	2 AAX91390_04	Continuation (5 of
c 43	154	19.1	273254	3 AAC81314	AAC81314 Chlamydia
44	152.5	18.9	4037	13 ADT05374	ADT05374 Haemophil
45	152.5	18.9	349980	13 ADT05648	ADT05648 Haemophil

ALIGNMENTS

RESULT 1
AAV63860
ID AAV63860 standard; DNA; 471 BP.
XX
AC AAV63860;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P45 (dUTPase) component DNA.
XX
KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication; ss.
XX
OS Pyrococcus furiosus; strain DSM 3638.
XX
FN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US005497.
XX
PR 21-MAR-1997; 97US-00822774.
PR 24-OCT-1997; 97US-00957709.
(STRA-) STRATAGENE.
PI Hogrefe H, Hansen CJ;
XX

DR WPI: 1998-542284/46.
DR P-PSDB; AAW72847.
XX Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX
XX
PS Claim 17; Page 43; 161pp; English.
XX
CC This DNA sequence encodes the P45 component (see AAW72847) of the
CC polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. PCR
CC primers (see AAW63861-64) based on an N-terminal peptide (see AAW72846)
CC of P45 were used to amplify P. furiosus genomic DNA, and sequencing was
CC carried out on purified PCR products and plasmid mini-preps. P45 and P40
CC (see AAW72844) are the predominant components of PEF, which acts to
CC enhance the activity of P. furiosus DNA polymerase, thereby providing
CC replication products of greater length and purity. P45 was identified as
CC a dUTPase, and possesses polymerase enhancing activity. The invention
CC provides novel extracts, proteins and complexes that improve the
CC polymerisation activity of nucleic acid polymerases, as well as DNA
CC constructs and antibodies. Also included are methods for identifying
CC compositions with polymerase enhancing activity, for purifying and using
CC these compositions, and specific extracts, proteins and complexes that
CC function to enhance polymerase activity. Nucleic acid polymerase
CC reactions can be enhanced (claimed) by mixing a nucleic acid template, at
CC least 1 polymerase and a composition having polymerase enhancing
CC activity. Kits are provided for replicating nucleic acids. The kits can
CC be used in site-directed mutagenesis, nucleic acid sequencing or
CC amplification (preferably PCR or RT-PCR). Isolated DNA can be used in the
CC recombinant production of P45. (Updated on 17-OCT-2003 to standardise OS
CC field)
XX
XX Sequence 471 BP; 156 A; 80 C; 124 G; 111 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.7e-97 Length: 471
Score: 806.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-08-957-709A-71 (1-156) x AAW63860 (1-471)

QY 1 MetLeuLeuProAspTrrpLysLeuGluLeuLeuGluProPheSerGluGlu 20
DB 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACCTTATAGAGCCATTTCAGAGAA 60
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
DB 61 TCGCTCCAAACAGCAGGTTATGCCTCAGAGTGGGAGAGAGGCTTTTGTAAAGGGGAAA 120
QY 41 LeuileAspValGluLysGluGlyLysValValileProArgGluTyrAlaLeuile 60
DB 121 TTAATCGACGTGGAAAGGAAAGGAAAGTGGTTATTCCTCCAAAGGGAATACGCCCTAATC 180
QY 61 LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleArgSer 80
DB 181 CTAACCCCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAGGAGC 240
QY 81 SerLeuAlaArgGluGlyValileGlySerPheAlaTrrpValAspProGlyTrrpAspGly 100
DB 241 AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGTGGTTGACCCAGGATGGATGGA 300
QY 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
DB 301 AACTTAACACTAATGCTCTACAATGCCTCAATGAACCTGTGCAATTAAGATATGAGAG 360
QY 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTrrpArgGly 140
DB 361 AGATTGTGCAGATCGCATTTTATAAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAGA 420
QY 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLysLeu 156

DB 421 AACTATCAGGGGAGCACAAAGGTAGCGTTTTCAGAGAGAAAGAACTC 468

RESULT 2

AAH41227
ID AAH41227 standard; DNA; 265118 BP.
XX AC AAH41227;
XX DT 29-OCT-2001 (first entry)
XX DE Pyrococcus abyssi genomic fragment #6.
XX KW Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX OS Pyrococcus abyssi.
XX FH Key Location/Qualifiers
FT misc_feature 1..49980
FT /tag= a
FT /note= "This sequence overlaps with the 3' end of
FT AAH41226"
XX FR2792651-Al.
XX PD 27-OCT-2000.
XX PF 21-APR-1999; 99FR-00005034.
XX PR 21-APR-1999; 99FR-00005034.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX (IFRE-) IFREMER INST FR RECH EXPL MER.
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX Querellou J, Weissenbach J, Saurin W, Heilig R;
XX WPI; 2001-126236/14.
XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins
XX useful in industry.
XX Claim 1; Page 593-665; 1657pp; French.
XX The present invention relates to the genomic sequence of *Pyrococcus*
XX abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
XX hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
XX vents. The present sequence is a fragment of the genomic sequence of P.
XX abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41226.
XX The proteins of the present invention have various potential industrial
XX uses, since the proteins are stable at very high temperatures, some up to
XX 110 degrees centigrade. Note: This patent is in the same patent family as
XX WO2000085082, which contains additional sequences as shown in AAB99132-
XX AAB99143, AAB75903-AAH75920 and AAG66436
XX Sequence 265118 BP; 75226 A; 61302 C; 54005 G; 74585 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 6.36e-80 Length: 265118
Score: 703.00 Matches: 131
Percent Similarity: 94.12% Conservative: 13
Best Local Similarity: 85.62% Mismatches: 9
Query Match: 87.22% Indels: 0
DB: 5 Gaps: 0
US-08-957-709A-71 (1-156) x AAH41227 (1-265118)
QY 1 MetLeuLeuProAspTrrpLysLeuGluLeuLeuGluProPheSerGluGlu 20
DB 262045 ATGCTCTCTCCAGACTGGAAATAGGAAGAGATTTTAAAGCAATTCACAGAGAA 262104
QY 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40

Db 262105 TCGCTTCAACAGCGCTGCGTACGACTTAAGGGTAGGCAAGGAGCTTATCATCTACGGGGAAG 262164

Qy 41 LeuileaspValGluLysGluGlyValValileProArgGluTyrAlaLeuille 60
 Db 262165 TTCATAGATGTTGAAGAAGGAGGGCAAGTCAATACTCTCAAAGAATAATGCCCTAATA 262224

Qy 61 LeuThrLeuGluArgLleLysLeuProAspAspValMetGlyAspMetLysLleArgSer 80
 Db 262225 CTGACCCCTAGAGAGGATAAAGCTTCCAGATGATATTATGGGAGACATGAAGATAAGGAGC 262284

Qy 81 SerLeuAlaArgGluGlyValLleGlySerPheAlaTrpValAspProGlyTrpAspGly 100
 Db 262285 AGTCTAGCTAGGAGAGGTGTTCTAGGCTCTCTTTCATGGGTAGACCCGGGATGGGACCGT 262344

Qy 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120
 Db 262345 AATCTAACCTTGATGCTTTCAATATGTCATCGAAAAAGGAGGTAAATTTTAAGGTACAAAGAG 262404

Qy 121 ArgPheValGlnLleAlaPheLleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140
 Db 262405 AGGTTTCTCAGATAGCGCTTCTTAAGGCTTCCGAGCTCCCGCAAGAAATCCATACAGGGGC 262464

Qy 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArg 153
 Db 262465 AACTATCAAGAGACAGAGATAGTCTCTCAAAGAGA 262503

RESULT 3
 ADN46845_12/C
 Continuation (13 of 21) of ADN46845 from base 1200001 (Thermococcus kodakaraensis KOD1 S
 WP Sequence split into 21 fragments LOCUS ADN46845 Accession Adn46845

Fragment Name	Begin	End
WP ADN46845_00	1	110000
WP ADN46845_01	100001	210000
WP ADN46845_02	200001	310000
WP ADN46845_03	300001	410000
WP ADN46845_04	400001	510000
WP ADN46845_05	500001	610000
WP ADN46845_06	600001	710000
WP ADN46845_07	700001	810000
WP ADN46845_08	800001	910000
WP ADN46845_09	900001	1010000
WP ADN46845_10	1000001	1110000
WP ADN46845_11	1100001	1210000
WP ADN46845_12	1200001	1310000
WP ADN46845_13	1300001	1410000
WP ADN46845_14	1400001	1510000
WP ADN46845_15	1500001	1610000
WP ADN46845_16	1600001	1710000
WP ADN46845_17	1700001	1810000
WP ADN46845_18	1800001	1910000
WP ADN46845_19	1900001	2010000
WP ADN46845_20	2000001	2089378

Alignment Scores:
 Pred. No.: 1.58e-79 Length: 110000
 Score: 696.00 Matches: 130
 Percent Similarity: 92.26% Conservative: 13
 Best Local Similarity: 83.87% Mismatches: 12
 Query Match: 86.35% Indels: 0
 DB: 12 Gaps: 0

US-08-957-709A-71 (1-156) x ADN46845_12 (1-110000)

Qy 1 MetLeuLeuProAspTrpLysLleArgLysGluLleLeuilleGluProPheSerGluGlu 20
 Db 40800 ATGATGCTTCCCGATTGGAGATTAGAAAGGAAATTTTGATCGAGCCTTTTCAGTGAAGAA 40741

Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40
 Db 40740 TCCCTGCAGCTGCAGGATATGACCTAAGGTTGGAGCGCAAGCTTACGTAATGGTAAA 40681

Qy 41 LeuilleaspValGluLysGluGlyValValileProArgGluTyrAlaLeuille 60
 Db 40740 TCCCTGCAGCTGCAGGATATGACCTAAGGTTGGAGCGCAAGCTTACGTAATGGTAAA 40681

Db	40680	ATACTCGACGTTAAGGATTTCAGGGGAGTTACAAATACCCCAAGACTTACGCTTGGTT	40680	ATCTCGACGTTAAGGATTTCAGGGGAGTTACAAATACCCCAAGACTTACGCTTGGTT	40680
Qy	61	LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleLeuArgSer	80	LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleLeuArgSer	80
Db	40620	TTGACTCTCGAGAGGATCAAGCTTCGCGATGATGTAATGGCGCATGAAGCTTAGGAGC	40560	TTGACTCTCGAGAGGATCAAGCTTCGCGATGATGTAATGGCGCATGAAGCTTAGGAGC	40560
Qy	81	SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly	100	SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly	100
Db	40560	AGTCTACGCCAGAGAGGGCTTAATCGGTTCCTTTGGCTGGGTTCACCTCGATGGATGGA	40500	AGTCTACGCCAGAGAGGGCTTAATCGGTTCCTTTGGCTGGGTTCACCTCGATGGATGGA	40500
Qy	101	AsnLeuThrLeuMetLeuTyTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu	120	AsnLeuThrLeuMetLeuTyTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu	120
Db	40500	AAATCTACTCTAGCCCTGTTCACAGCTTCACACGAGTCTGTGGACATCAATACGGCGAG	40440	AAATCTACTCTAGCCCTGTTCACAGCTTCACACGAGTCTGTGGACATCAATACGGCGAG	40440
Qy	121	ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly	140	ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly	140
Db	40440	CGCTTTGTTTCAGATAGCGTTTATCCGGCTGGAGGGCGCTCTAAAAACCCCTATCGTGA	40380	CGCTTTGTTTCAGATAGCGTTTATCCGGCTGGAGGGCGCTCTAAAAACCCCTATCGTGA	40380
Qy	141	AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys	155	AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys	155
Db	40380	AAATACCGAGGAAGCAACATCTAGCGCTCTCAAAAGGAACCG	40336	AAATACCGAGGAAGCAACATCTAGCGCTCTCAAAAGGAACCG	40336
RESULT 4					
ADN47591_08					
Continuation (9 of 21) of ADN47591 from base 800001 (Thermococcus kodakaraensis					
WP Sequence split into 21 fragments LOCUS ADN47591 Accession Adn47591					
WP	Fragment Name	Begin	End		
WP	ADN47591_00	1	110000		
WP	ADN47591_01	100001	210000		
WP	ADN47591_02	200001	310000		
WP	ADN47591_03	300001	410000		
WP	ADN47591_04	400001	510000		
WP	ADN47591_05	500001	610000		
WP	ADN47591_06	600001	710000		
WP	ADN47591_07	700001	810000		
WP	ADN47591_08	800001	910000		
WP	ADN47591_09	900001	1010000		
WP	ADN47591_10	1000001	1110000		
WP	ADN47591_11	1100001	1210000		
WP	ADN47591_12	1200001	1310000		
WP	ADN47591_13	1300001	1410000		
WP	ADN47591_14	1400001	1510000		
WP	ADN47591_15	1500001	1610000		
WP	ADN47591_16	1600001	1710000		
WP	ADN47591_17	1700001	1810000		
WP	ADN47591_18	1800001	1910000		
WP	ADN47591_19	1900001	2010000		
WP	ADN47591_20	2000001	2089378		
Alignment Scores:					
Pred. No.:	1.58e-79	Length:	110000		
Score:	696.00	Matches:	130		
Percent Similarity:	92.26%	Conservative:	13		
Best Local Similarity:	83.87%	Mismatches:	12		
Query Match:	86.35%	Indels:	0		
DB:	12	Gaps:	0		
US-08-957-709A-71 (1-156) x ADN47591_08 (1-110000)					
Qy	1	MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu	20	MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu	20
Db	48578	ATGATGTTCCCGATTGGAAGATTAGAAAAGAAATTTTGATCGAGCGCTTTTCAGTGAGAA	4863	ATGATGTTCCCGATTGGAAGATTAGAAAAGAAATTTTGATCGAGCGCTTTTCAGTGAGAA	4863
Qy	21	SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys	40	SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys	40
Db	48638	TCCTCGAGCGCTCGAGATTATGACCTTAGGGTTGGAGCGAGACTTACGTAATAGTAA	4869	TCCTCGAGCGCTCGAGATTATGACCTTAGGGTTGGAGCGAGACTTACGTAATAGTAA	4869
Qy	41	LeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyrAlaLeuIle	60	LeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyrAlaLeuIle	60
Db	48698	ATACTCGACGTTAAGGATTTCAGGGGAGTTTACAATACCCCAAGAACTTACGCTTGGTT	4875	ATACTCGACGTTAAGGATTTCAGGGGAGTTTACAATACCCCAAGAACTTACGCTTGGTT	4875
Qy	61	LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleLeuArgSer	80	LeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIleLeuArgSer	80

Db	48758	TTGACTCTCGAGAGGATCAAGCTTCGGATGATGTAATAGGGCGACATGAAGCTTAGGAGC	48817
Qy	81	SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly	100
Db	48818	AGCTAGGCAGAGAGGCTTAATCGGTTCTTTTGGCTGGGTTCAGCCCTGGATGGATGGA	48877
Qy	101	AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu	120
Db	48878	AATCTCACTCTAGCCCTGTTCAACGCTTCCAACGAGTCTGTGGAACTCAAATACGGCGAG	48937
Qy	121	ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly	140
Db	48938	CGCTTTGTTTCAGATAGCGTTTATCCGGCTGGAGGGCCCTGCTAAAAACCCCTATCGTGA	48997
Qy	141	AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys	155
Db	48998	AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAAGGAAACGG	49042
RESULT 5			
Continuation (13 of 21) of ADN46123 from base 1200001 (Thermococcus kodakaraensis KOD1 9			
WP Sequence split into 21 fragments LOCUS ADN46123 Accession Adn46123			
WP	Fragment Name	Begin	End
WP	ADN46123_00	1	110000
WP	ADN46123_01	100001	210000
WP	ADN46123_02	200001	310000
WP	ADN46123_03	300001	410000
WP	ADN46123_04	400001	510000
WP	ADN46123_05	500001	610000
WP	ADN46123_06	600001	710000
WP	ADN46123_07	700001	810000
WP	ADN46123_08	800001	910000
WP	ADN46123_09	900001	1010000
WP	ADN46123_10	1000001	1110000
WP	ADN46123_11	1100001	1210000
WP	ADN46123_12	1200001	1310000
WP	ADN46123_13	1300001	1410000
WP	ADN46123_14	1400001	1510000
WP	ADN46123_15	1500001	1610000
WP	ADN46123_16	1600001	1710000
WP	ADN46123_17	1700001	1810000
WP	ADN46123_18	1800001	1910000
WP	ADN46123_19	1900001	2010000
WP	ADN46123_20	2000001	2089378
Alignment Scores:			
Pred. No.:	1.58e-79	Length:	110000
Score:	696.00	Matches:	130
Percent Similarity:	92.26%	Conservative:	13
Best Local Similarity:	83.87%	Mismatches:	12
Query Match:	86.35%	Indels:	0
DB:	12	Gaps:	0
US-08-957-709a-71 (1-156) x ADN46123_12 (1-110000)			
Qy	1	MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu	20
Db	40800	ATGATGCTTCCCGATTGGAAGATTAGAAAGGAAATTTTTCAGTCAGCCCTTTCAGTGAAGAA	40741
Qy	21	SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys	40
Db	40740	TCCTGCGACCTTCAGAGATTACCTAAGGGTTGGAGGCGAAGCTTACGTAATCGTAAA	40681
Qy	41	LeuIleAspValGluLysGluGlyValValIleProProArgGluTyrAlaLeuIle	60
Db	40680	ATACTCGAGCTTAAGGATTTCAGGGGAGTTTACAATACCCCCAAGACTTACGCCCTTGTT	40621
Qy	61	LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer	80
Db	40620	TTGACTCTCGAGAGGATCAAGCTTCGGATGATGTAATAGGGCGACATGAAGCTTAGGAGC	40561
Qy	81	SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly	100

Db	40560	AGTCTAGCCAGAGAGGCTTAATCGGTTCTTTTGGCTGGGTTCAGCCCTGGATGGATGGA	40501
Qy	101	AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu	120
Db	40500	AATCTCACTCTAGCCCTGTTTCAACGCTTCCAACGAGTCTGTGGAACTCAAATACGGCGAG	40441
Qy	121	ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly	140
Db	40440	CGCTTTGTTTCAGATAGCGTTTATCCGGCTGGAGGGCCCTGCTAAAAACCCCTATCGTGA	40381
Qy	141	AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys	155
Db	40380	AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAAGGAAACGG	40336
RESULT 6			
Continuation (9 of 21) of ADN47209 from base 800001 (Thermococcus kodakaraensis KOD1 9er			
WP Sequence split into 21 fragments LOCUS ADN47209 Accession Adn47209			
WP	Fragment Name	Begin	End
WP	ADN47209_00	1	110000
WP	ADN47209_01	100001	210000
WP	ADN47209_02	200001	310000
WP	ADN47209_03	300001	410000
WP	ADN47209_04	400001	510000
WP	ADN47209_05	500001	610000
WP	ADN47209_06	600001	710000
WP	ADN47209_07	700001	810000
WP	ADN47209_08	800001	910000
WP	ADN47209_09	900001	1010000
WP	ADN47209_10	1000001	1110000
WP	ADN47209_11	1100001	1210000
WP	ADN47209_12	1200001	1310000
WP	ADN47209_13	1300001	1410000
WP	ADN47209_14	1400001	1510000
WP	ADN47209_15	1500001	1610000
WP	ADN47209_16	1600001	1710000
WP	ADN47209_17	1700001	1810000
WP	ADN47209_18	1800001	1910000
WP	ADN47209_19	1900001	2010000
WP	ADN47209_20	2000001	2089378
Alignment Scores:			
Pred. No.:	1.58e-79	Length:	110000
Score:	696.00	Matches:	130
Percent Similarity:	92.26%	Conservative:	13
Best Local Similarity:	83.87%	Mismatches:	12
Query Match:	86.35%	Indels:	0
DB:	12	Gaps:	0
US-08-957-709a-71 (1-156) x ADN47209_08 (1-110000)			
Qy	1	MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu	20
Db	48578	ATGATGCTTCCCGATTGGAAGATTAGAAAGGAAATTTTTCAGTCAGCCCTTTCAGTGAAGAA	48637
Qy	21	SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys	40
Db	48638	TCCTGCGACCTTCAGAGATTACCTAAGGGTTGGAGGCGAAGCTTACGTAATCGTAAA	48697
Qy	41	LeuIleAspValGluLysGluGlyValValIleProProArgGluTyrAlaLeuIle	60
Db	48698	ATACTCGAGCTTAAGGATTTCAGGGGAGTTTACAATACCCCCAAGACTTACGCCCTTGTT	48757
Qy	61	LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer	80
Db	48758	TTGACTCTCGAGAGGATCAAGCTTCGGATGATGTAATAGGGCGACATGAAGCTTAGGAGC	48817
Qy	81	SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly	100
Db	48818	AGTCTAGCCAGAGAGGCTTAATCGGTTCTTTTGGCTGGGTTCAGCCCTGGATGGATGGA	48877
Qy	101	AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu	120

Db 48878 AATCTCACTCTAGCCCTGTTCAACGCTTCCACGAGTCTGTGGAACTCAATACGGCGAG 48937

Qy 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140

Db 48938 CGCTTTGTTTCAGATAGCGTTTATCCGCTCGAGGGCGCTGCTAAATAACCCCTATCGTGA 48997

Qy 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys 155

Db 48998 AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAGGAACCG 49042

RESULT 7

ADN46464_12/c
Continuation (13 of 21) of ADN46464 from base 1200001 (Thermococcus kodakaraensis KOD1
WP Sequence split into 21 fragments LOCUS ADN46464 Accession ADN46464

Fragment Name	Begin	End
WP ADN46464_00	1	110000
WP ADN46464_01	100001	210000
WP ADN46464_02	200001	310000
WP ADN46464_03	300001	410000
WP ADN46464_04	400001	510000
WP ADN46464_05	500001	610000
WP ADN46464_06	600001	710000
WP ADN46464_07	700001	810000
WP ADN46464_08	800001	910000
WP ADN46464_09	900001	1010000
WP ADN46464_10	1000001	1110000
WP ADN46464_11	1100001	1210000
WP ADN46464_12	1200001	1310000
WP ADN46464_13	1300001	1410000
WP ADN46464_14	1400001	1510000
WP ADN46464_15	1500001	1610000
WP ADN46464_16	1600001	1710000
WP ADN46464_17	1700001	1810000
WP ADN46464_18	1800001	1910000
WP ADN46464_19	1900001	2010000
WP ADN46464_20	2000001	2089378

Alignment Scores:
Pred. No.: 1.58e-79 Length: 110000
Score: 696.00 Matches: 130
Percent Similarity: 92.26% Conservative: 13
Best Local Similarity: 83.87% Mismatches: 12
Query Match: 86.35% Indels: 0
DB: 12 Gaps: 0

US-08-957-709a-71 (1-156) x ADN46464_12 (1-110000)

Qy 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu 20

Db 40800 ATGATGCTTCCCGATTGGAGATTAGAAAGGAATTTTGATCGAGCCTTTCAGTGAAGAA 40741

Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40

Db 40740 TCCTCGAGCTTCAGGATATGACCTTAGGGTTGGAGGCGAAGCTTACGTAATGGTAAA 40681

Qy 41 LeuIleAspValGluLysGluLysValValIleProProArgGluTyrAlaLeuIle 60

Db 40680 ATACTCGAGTTAAGGATTCAGGGGAGTTTACAATACCCCAAGACTTACGCCCTTGGTT 40621

Qy 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80

Db 40620 TTGACTCTCGAGAGGATCAAGCTTCCGGATGATGTAATGGCGACATGAAGCTTAGGAGC 40561

Qy 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly 100

Db 40560 AGTCTAGCCAGAGGGCTTAATCGGTTCTTTTGGCTGGGTTGACCTGGATGGGATGA 40501

Qy 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120

Db 40500 AATCTCACTAGCCCTGTTCAACGCTTCCACGAGTCTGTGGAACTCAAAATACGGCGAG 40441

Qy 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140

Db 40440 CGCTTTGTTTCAGATAGCGTTTATCCGCTGGAGGGCGCTGCTAAAAACCCCTATCGTGA 40381

Qy 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys 155

Db 40380 AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAGGAACCG 40336

RESULT 8

ADN47960_08
Continuation (9 of 21) of ADN47960 from base 800001 (Thermococcus kodakaraensis KOD1 ge
WP Sequence split into 21 fragments LOCUS ADN47960 Accession ADN47960

Fragment Name	Begin	End
WP ADN47960_00	1	110000
WP ADN47960_01	100001	210000
WP ADN47960_02	200001	310000
WP ADN47960_03	300001	410000
WP ADN47960_04	400001	510000
WP ADN47960_05	500001	610000
WP ADN47960_06	600001	710000
WP ADN47960_07	700001	810000
WP ADN47960_08	800001	910000
WP ADN47960_09	900001	1010000
WP ADN47960_10	1000001	1110000
WP ADN47960_11	1100001	1210000
WP ADN47960_12	1200001	1310000
WP ADN47960_13	1300001	1410000
WP ADN47960_14	1400001	1510000
WP ADN47960_15	1500001	1610000
WP ADN47960_16	1600001	1710000
WP ADN47960_17	1700001	1810000
WP ADN47960_18	1800001	1910000
WP ADN47960_19	1900001	2010000
WP ADN47960_20	2000001	2089378

Alignment Scores:
Pred. No.: 1.58e-79 Length: 110000
Score: 696.00 Matches: 130
Percent Similarity: 92.26% Conservative: 13
Best Local Similarity: 83.87% Mismatches: 12
Query Match: 86.35% Indels: 0
DB: 12 Gaps: 0

US-08-957-709a-71 (1-156) x ADN47960_08 (1-110000)

Qy 1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuIleGluProPheSerGluGlu 20

Db 48578 ATGATGCTTCCCGATTGGAGATTAGAAAGGAATTTTGATCGAGCCTTTCAGTGAAGAA 48637

Qy 21 SerLeuGlnProAlaGlyTyrAspLeuArgValGlyArgGluAlaPheValLysGlyLys 40

Db 48638 TCCTCGAGCTTCAGGATATGACCTTAGGGTTGGAGGCGAAGCTTACGTAATGGTAAA 48697

Qy 41 LeuIleAspValGluLysGluLysValValIleProProArgGluTyrAlaLeuIle 60

Db 48698 ATACTCGAGTTAAGGATTCAGGGGAGTTTACAATACCCCAAGACTTACGCCCTTGGTT 48757

Qy 61 LeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIleArgSer 80

Db 48758 TTGACTCTCGAGAGGATCAAGCTTCCGGATGATGTAATGGCGACATGAAGCTTAGGAGC 48817

Qy 81 SerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTyrAspGly 100

Db 48818 AGTCTAGCCAGAGGGCTTAATCGGTTCTTTTGGCTGGGTTGACCTGGATGGGATGA 48877

Qy 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGlu 120

Db 48878 AATCTCACTAGCCCTGTTCAACGCTTCCACGAGTCTGTGGAACTCAAAATACGGCGAG 48937

Qy 121 ArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArgGly 140

Db 48938 CGCTTTGTTTCAGATAGCGTTTATCCGCTGGAGGGCGCTGCTAAAAACCCCTATCGTGA 48997

Qy 141 AsnTyrGlnGlySerThrArgLeuAlaPheSerLysArgLysLys 155

```

Db 48998 AATTACAGGAAGCAACATCTAGCGCTCTCAAAAAGGAACGG 49042
RESULT 9
AAQ62300/c
ID AAQ62300 standard; DNA; 3382 BP.
XX
XX AAQ62300;
XX
XX 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 27-MAY-1994 (first entry)
XX
XX pDam-L3 plasmid fragment encoding thermostable ligase.
XX
XX Ligase, thermostable; thermostability; Desulfurolobus ambivalens;
KW archaeobacteria; detection; ligation; ATP; adenosine triphosphate; ss.
XX
XX Acidianus ambivalens.
XX
XX Key Location/Qualifiers
FH 1..273
FT CDS /tag= a
FT /note= "Open reading frame 4."
FT complement(250..771)
FT /tag= b
FT /note= "Open reading frame 3."
FT 878..883
FT promoter /tag= c
FT 908..2707
FT CDS /tag= d
FT /product= "Ligase."
FT 2713..2722
FT terminator /tag= e
FT /note= "Potential transcription termination signal."
FT 2804..2811
FT terminator /tag= f
FT /note= "Experimental transcription termination signal."
FT 2961..3382
FT CDS /tag= g
FT /note= "Open reading frame 2."
FT
FT
XX DE4217134-A1.
XX
XX 25-NOV-1993.
XX
XX 23-MAY-1992; 92DE-04217134.
XX
XX 23-MAY-1992; 92DE-04217134.
XX
XX (BOE ) BOEHRINGER MANNHEIM GMBH.
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Rueger R, Kessler C, Kaletta C, Jarsch M, Kletzin A;
XX
XX WPI: 1993-378402/48.
DR P-PSDB; AAR44614, AAR51076, AAR51077, AAR51078.
XX
XX Thermostable ligase from archaeobacteria - and DNA coding for it, useful
PT for nucleic acid detection.
XX
XX Claim 11; Fig 1; 26pp; German.
XX
XX The DNA ligase is ATP dependent and is useful for detecting nucleic acids
CC by hybridising two oligonucleotides with adjacent sequences of the target
CC nucleic acid, ligating the two oligonucleotides (using the ligase) and
CC detecting the ligation product. (Updated on 25-MAR-2003 to correct PN
CC field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
XX Sequence 3382 BP; 1188 A; 512 C; 683 G; 999 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 3,28e-21 Length: 3382
Score: 242.00 Matches: 58

```

```

Percent Similarity: 57.14% Conservative: 30
Best Local Similarity: 37.66% Mismatches: 52
Query Match: 30.02% Indels: 15
Db: 2 Gaps: 3
US-08-957-709A-71 (1-156) x AAQ62300 (1-3382)
QY 12 IleLeuIleGluProPheSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgVal 31
Db 723 ATAGTAATCTCTCTCTTACTCAGATACTATAGAGAGACGGGGTGCATTTTAGAGTA 664
QY 32 Gly-----ArgGluAlaPheValLysGlyLys----- 40
Db 663 GGAGGAGAGATAGCCCGCTTTTAAGAAGACTGACGAATATATAGAGATGGAAGACCCG 604
QY 41 -----LeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyrAla 58
Db 603 CGTTCAATTTATGAGATAGAGAAAGAGATGAATTCATCATTTACCCCTAACCAACATGTA 544
QY 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIle 78
Db 543 CTGTTAGTTACTGAGGAGTATGTTAACTGCGAAGACGTAATGGCTTCGTCACCTTA 484
QY 79 ArgSerSerLeuAlaArgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrp 98
Db 483 AGGCTCTCTTTGCCAGGCTGGTCTTTTGTACCACCGACTATCGTAGATCGAGGCTTT 424
QY 99 AspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyr 118
Db 423 GAAGGACAACTCACAATAGAAAGTACTA---GGTTCTGCTTTTCCCGTGAAGATAAGAGG 367
QY 119 GlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyr 138
Db 366 GGGACTCGCTTTCTTCACCTCATTTTCGCCAGACATTAACACCACTAGAGATCCTTAT 307
QY 139 ArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 152
Db 306 CATGCAAAATACCAAGGACA-ACAAGGGGTAACTTTTACCAA 266
RESULT 10
ADK16664
ID ADK16664 standard; DNA; 597 BP.
XX
XX AC ADK16664;
XX
XX 06-MAY-2004 (first entry)
XX
XX Nanoarchaeum equitans cancer-associated (CA) gene #308.
XX cancer-associated gene; CA gene; cancer; carcinoma; lymphoma; leukaemia;
KW ds; gene.
XX
XX Nanoarchaeum equitans.
XX
XX WO2003093434-A2.
XX
XX 13-NOV-2003.
XX
XX 01-MAY-2003; 2003WO-US013699.
XX
XX 01-MAY-2002; 2002US-0377447P.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Stetter KO, Waters E, Kretz K, Podar M, Richardson T;
PI Noordewier M;
XX
XX WPI: 2004-053041/05.
DR P-PSDB; ADK16665.
XX
XX New recombinant cancer-associated genes, such as KCNJ9, useful for
PT diagnosing or treating carcinoma, e.g. breast, colon, rectal, pancreatic,
PT cervical, or skin cancers, lymphomas, or leukemia.

```


XX Claim 5; SEQ ID NO 616; 251pp; English.
XX The invention comprises then amino acid and coding sequences of cancer-associated (CA) genes isolated from Nanoarchaeum equitans. The invention CC also comprises the Nanoarchaeum equitans genome. The DNA and protein CC sequences of the invention are useful for diagnosing and treating cancer CC (e.g. carcinoma, lymphoma, or leukaemia). The present DNA sequence CC represents a Nanoarchaeum equitans CA gene of the invention.
XX
SQ Sequence 597 BP; 246 A; 80 C; 104 G; 167 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,45e-20 Length: 597
Score: 227.50 Matches: 63
Percent Similarity: 49.47% Conservative: 31
Best Local Similarity: 33.16% Mismatches: 51
Query Match: 28.23% Indels: 45
DB: 12 Gaps: 8

US-08-957-709A-71 (1-156) x ADK16664 (1-597)

QY 1 MetLeuLeuProAspTrpLysIle-----ArgLysGluIleLeuIleGlu 15
DB 1 ATGCTATTAAATGATAGAGAAATTAAGAGCTAATAGACAAAAGAGATAATATTAG 60
QY 16 ProPheSerGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
DB 61 CCTTTTTCATGAACAAATCCACCAGCATCGATAGATCTAAGGCTAGGCAATAGATT 120
QY 34 GluAlaPheValLysGly-----LysLeu 41
DB 121 AGAATTTTCAGAAAAGCGGAAATAGAGTAAATTGATCCCAAGGATTTTAAGGATGA 180
QY 42 IleAspValGluLys----- 46
DB 181 ATAAGATAGACACAGATGAAATAAATAAATAATTTGAAAAATACAAATATACCGATG 240
QY 47 -----GluGlyLysValValIleProProArgGluTyrAlaLeuIleLeuThrLeuGlu 64
DB 241 ATTACAGAGAACCTTTTCATTATTATCCAGGGGATTTTTGTAGCCTCTATATATGAA 300
QY 65 ArgIleLeuProAspAspValMetGlyAspMetLysIleArgSerSerLeuAlaArg 84
DB 301 TATATAAAATTCGCAAGATATATAGCAGCACAAATTCATGGCAAAATCCTCTATAG 360
QY 85 GluGlyValIle-----GlySerPheAlaTtpValaspProGlyTrpaspGlyAsnLeu 102
DB 361 TTAGGTTTAAATATACATACATCTGCAGGTGGATAGACCCCTGGTTATGAGGCCATCTA 420
QY 103 ThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGluArgPhe 122
DB 421 ACATTAGAATAATTTAATACTACAAATGCGCAATTAAGCTTTACCTAAAATGAAAAA 480
QY 123 ValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArg----- 139
DB 481 GCACAAATTCGAGCTATTAGAATA---AATCCTGTAGAGAGAGACTATAAAGAAAAAGGA 537
QY 140 GlyAsnTyr-----GlnGlySerThr 146
DB 538 GGTAATACTATAAAGAAAAGGTGCTACC 567

RESULT 11

ADK16049_2
Continuation (3 of 5) of ADK16049 from base 200001 (Nanoarchaeum equitans genome.)
WP Sequence split into 5 fragments LOCUS ADK16049 Accession ADK16049

WP	Fragment Name	Begin	End
WP	ADK16049_1	1	110000
WP	ADK16049_2	100001	210000
WP	ADK16049_3	200001	310000
WP	ADK16049_4	300001	410000
WP	ADK16049_5	400001	490885

Alignment Scores:
Pred. No.: 3,71e-17 Length: 110000
Score: 227.50 Matches: 63
Percent Similarity: 49.47% Conservative: 31
Best Local Similarity: 33.16% Mismatches: 51
Query Match: 28.23% Indels: 45
DB: 12 Gaps: 8

US-08-957-709A-71 (1-156) x ADK16049_2 (1-110000)

QY 1 MetLeuLeuProAspTrpLysIle-----ArgLysGluIleLeuIleGlu 15
DB 83658 ATGCTATTAAATGATAGAGAAATTAAGAGCTAATAGACAAAAGAGATAATATTAG 83717
QY 16 ProPheSerGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
DB 83718 CCTTTTTCATGAACAAATCCACCAGCATCGATAGATCTAAGGCTAGGCAATAGATT 83777
QY 34 GluAlaPheValLysGly-----LysLeu 41
DB 83778 AGAATTTTCAGAAAAGCGGAAATAGAGTAAATTGATCCCAAGGATTTTAAGGATGA 83837
QY 42 IleAspValGluLys----- 46
DB 83838 ATAAGATAGACACAGATGAAATAAATAAATAATTTGAAAAATACAAATATACCGATG 83897
QY 47 -----GluGlyLysValValIleProProArgGluTyrAlaLeuIleLeuThrLeuGlu 64
DB 83898 ATTACAGAGAACCTTTTCATTATTATCCAGGGGATTTTTGTAGCCTCTATATATGAA 83957
QY 65 ArgIleLeuProAspAspValMetGlyAspMetLysIleArgSerSerLeuAlaArg 84
DB 83958 TATATAAAATTCGCAAGATATATAGCAGCACAAATTCATGGCAAAATCCTCTATAG 84017
QY 85 GluGlyValIle-----GlySerPheAlaTtpValaspProGlyTrpaspGlyAsnLeu 102
DB 84018 TTAGGTTTAAATATACATACATCTGCAGGTGGATAGACCCCTGGTTATGAGGCCATCTA 84077
QY 103 ThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGlyGluArgPhe 122
DB 84078 ACATTAGAATAATTTAATACTACAAATGCGCAATTAAGCTTTACCTAAAATGAAAAA 84137
QY 123 ValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsnProTyrArg----- 139
DB 84138 GCACAAATTCGAGCTATTAGAATA---AATCCTGTAGAGAGAGACTATAAAGAAAAAGGA 84194
QY 140 GlyAsnTyr-----GlnGlySerThr 146
DB 84195 GGTAATACTATAAAGAAAAGGTGCTACC 84224

RESULT 12

ADK27081_02/c

Continuation (3 of 17) of ADM27081 from base 200001 (Hyperthermophile Methanopyrus kandi)
WP Sequence split into 17 fragments LOCUS ADM27081 Accession ADM27081

WP	Fragment Name	Begin	End
WP	ADM27081_00	1	110000
WP	ADM27081_01	100001	210000
WP	ADM27081_02	200001	310000
WP	ADM27081_03	300001	410000
WP	ADM27081_04	400001	510000
WP	ADM27081_05	500001	610000
WP	ADM27081_06	600001	710000
WP	ADM27081_07	700001	810000
WP	ADM27081_08	800001	910000
WP	ADM27081_09	900001	1010000
WP	ADM27081_10	1000001	1110000
WP	ADM27081_11	1100001	1210000
WP	ADM27081_12	1200001	1310000
WP	ADM27081_13	1300001	1410000
WP	ADM27081_14	1400001	1510000
WP	ADM27081_15	1500001	1610000
WP	ADM27081_16	1600001	1694968

03-AUG-2000; 2000JP-00280988.

(KYOW) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
Tateishi N, Senoh A, Ikeda M, Ozaki A;
WPI; 2001-376931/40.
P-PSDB; AAG92886.

Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analyzing expression profile or pattern of a gene and identifying homologous gene.

Claim 8; SEQ ID NO 3140; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office

Sequence 567 BP; 104 A; 132 C; 165 G; 166 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.43e-18 Length: 567
Score: 214.00 Matches: 56
Percent Similarity: 50.30% Conservative: 29
Best Local Similarity: 33.14% Mismatches: 60
Query Match: 26.55% Indels: 24
DB: 5 Gaps: 4

US-08-957-709A-71 (1-156) x AAH68105 (1-567)

Qy 1 MetLeuLeuProAspTrpLysIleArgLys-----GluIleLeuIleGlu 15
Db 1 GTGCTTCTTCAGATCGTGACATCGTAATCAATTGACGCGGCGACTTGGGAATTGAA 60
Qy 16 ProPheSerGluGluSerLeuGluInProAlaGlyTyrAspLeuArgValGlyArg----- 33
Db 61 CCTTTCACGCTGAGCTGATTTCAGCCGTCGAGTGTGCATGTCGCGATCGACCGCTACTTTC 120
Qy 34 GluAlaPheValLysGlyLys----- 40
Db 121 CGGGTTTTCATAACTCTTAAGTACACCCACATGACCTAAGTTGAATCAGATGAGCTG 180
Qy 41 -----LeuIleAspValGluLysGluGlyLysValValIleProArgGluTyrAla 58
Db 181 ACAGCCCTTGTGAGGTGTGAGGACGGCGGAGCGGCTTGTGTCATCGGGTGAGTTGTG 240
Qy 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspValMetGlyAspMetLysIle 78
Db 241 CTGGCGTCCAGCTGGAAAAGTTTCACTTTTGCTGCGCATCTGGCTGGTTCGTTGGAGGCT 300
Qy 79 ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
Db 301 AAGTCGCTCTTGTGCTGCTTGGGTGTGTGACGCACTCTACTGCTGGTGTTCATTGATCCT 360
Qy 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db 361 GGTTTGTAGTGTTCATACACAGCTTGAGTGTGCCATGTGGCTAACTGCCGATCACGTTG 420
Qy 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
Db 421 TGCCCGGCGATGAGGTGGGCGAGTTGGCTTTGTTCAGATGAGTGTCCCTCGGAGACT 480

CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids, described
CC particularly L-lysine. The present sequence is a nucleic acid, described
CC in the exemplification of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the European Patent Office
XX
SQ Sequence 349980 BP; 81250 A; 97718 C; 90621 G; 80391 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.18e-14 Length: 349980
Score: 214.00 Matches: 56
Percent Similarity: 50.30% Conservative: 29
Best Local Similarity: 33.14% Mismatches: 60
Query Match: 26.55% Indels: 24
DB: 5 Gaps: 4

US-08-957-709A-71 (1-156) x AAH68533 (1-349980)

QY 1 MetLeuLeuProAspTrpLysIleArgLys-----GluIleLeuIleGlu 15
Db 337411 GTGCTTCTTCAGATCGTGACATCGTAATCAATTGACGCGGCGACTTGGGAATTGAA 337352
QY 16 ProPheSerGluGluSerLeuLeuGlnProAlaGlyTyrAspLeuArgValGlyArg----- 33
Db 337351 CCTTTCGACGCTGAGCTGATTCCGCCGTCGAGTGTGATGTCGCCGATGACGCGCTACTTC 337292
QY 34 GluAlaPheValLysGlyLys----- 40
Db 337291 CGGGTTTTCATAACTCTAAGTACACCCACCATTTGACCTTAAGTTGAATCAGGATGAGCTG 337232
QY 41 -----LeuIleAspValGluLysGluGlyLysValValIleProProArgGluTyrAla 58
Db 337231 ACCAGCCTTGTGAGGTTGAGGACGCGGAGGGCTTTGTGCTGCATCCGGGTGAGTTGTG 337172
QY 59 LeuIleLeuThrLeuGluArgIleLysLeuProAspAspValMetGlyAspMetLysIle 78
Db 337171 CTGGCGTCCACCGCTGGAAAGTTCACTTTCCTGCGCATCTGGCTGGTGGTTCGTTGGAGGGT 337112
QY 79 ArgSerSerLeuAlaArgGluGlyValIle-----GlySerPheAlaTrpValAspPro 96
Db 337111 AAGTCGTCTCTTGGTCTGCTTGGCTTGTGACGCACTCTACTGCTGGTTCATTCATCCT 337052
QY 97 GlyTrpAspGlyAsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 116
Db 337051 GGTTTTAGTGGTTACATCAGTTGGAGTTGTCCAATGTGGCTAATCTGCCGATCAGTTG 336992
QY 117 ArgTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArgAsn 136
Db 336991 TGGCCGGGTATGAAGTGGGGGAGTGGCTTTGTTCCAGATGAGTTCCTCCGAGACT 336932
QY 137 ProTyrArgGlyAsnTyrGlnGlySer 145
Db 336931 CCGTATGGTTCGCGCAAGCTTGGTTTCG 336905

Search completed: July 2, 2005, 09:23:30
Job time : 763 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 06:11:51 ; Search time 2526 Seconds
(without alignments)
7097.494 Million cell updates/sec

Title: US-08-957-709A-70

Perfect score: 471

Sequence: 1 atgctacttcagactggaa.....caagagagaagaactcttag 471

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gse1:*

9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41.6	8.8	712	4	BJ624188
2	40.4	8.6	1101	9	CNS0039G
3	39.6	8.4	756	9	AG530407 Mus muscu
4	39	8.3	336	2	BF281615 EST446206
5	38	8.1	714	9	AG318648 Mus muscu
6	37.8	8.0	1101	9	CNS0106X
7	37.6	8.0	604	9	CW017266 ZMMBLC001
8	37.6	8.0	808	9	AG404615 Mus muscu
9	37.6	8.0	1101	9	CNS0006J
10	37	7.9	1019	9	CNS0107N
11	36.8	7.8	997	9	CNS005TE
12	36.6	7.8	409	9	CE439533 tigr-gss-
13	36.4	7.7	577	8	AQ416995 RFI-11-1
14	36.2	7.7	841	7	CK196609 FGAS00506
15	36	7.6	519	8	BZ915075 CH240 57D
16	36	7.6	522	4	BI501756 tnm04f06.y
17	36	7.6	604	4	BJ106515
18	36	7.6	613	4	BJ770456 BJ770456
19	36	7.6	661	7	CK580840 IST.W15.2
20	36	7.6	893	5	BQ139899 NF026C1IP
21	36	7.6	973	4	BG261309 602373061
22	35.8	7.6	703	9	AG323494 Mus muscu
23	35.8	7.6	1101	9	AL078714 Drosophil
24	35.6	7.6	857	9	CR247912 Forward s

RESULT 1

BJ624188

LOCUS

DEFINITION

BJ624188

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Xenopus laevis

Db 395 TGGAGTGGATGGTATTATGAGCAGGAGGGGAGGAGGATAGGGGGGGTTCCTAA 336
Qy 278 GGGTTGACCCAGGATGGATGGAACCTTAACACCTAATGCTCTACAAATGCCTCAAAATGAAC 337
Db 335 GGGTAAACACAGGAAGGGGATACATTTAAATGTAATAACGAAATATCTAATAAAAA 276
Qy 338 CTGTGCAATTAAGATATGGAGAGATTTGTGCAGATCGCATTTAT 383
Db 275 ATGTAGATTTTCAGATATGGGAAATTAGATAGTGAATTTCTTTGT 230

RESULT 4
BF281615/c
LOCUS
DEFINITION
EST446206 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
Rattus norvegicus cDNA clone RGIAN54, mRNA sequence.

ACCESSION
BF281615
VERSION
BF281615.1 GI:11212685
SOURCE
EST.
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 336)
Malek, R.L., Cho, J., Lee, Y., Karamycheva, S., Parvizi, B., Pertea, G.,
Sultana, R., Tsai, J., White, J., Quackenbush, J. and Lee, N.H.
Generation of ESTs from Normalized Rat Embryo, Bento Soares
Unpublished (2000)
Other ESTs: EST446205 EST446207
Contact: Lee, NH

LOCUS
DEFINITION
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.

FEATURES
source
1..336
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="RGIAN54"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/notes="Vector: pT377pac; Site 1: EcoRI; Site 2: NotI;
Combination of ROV, RBR, RKI, RLI, RPL, RLU, REM, RMI,
RSP, RHE, RPC, RPN"

ORIGIN
Query Match 8.3%; Score 39; DB 2; Length 336;
Best Local Similarity 52.1%; Pred. No. 2;
Matches 87; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 90 AGTGGCAGAGAGCGTTTCTTAAAGGGAATTAATCGACGTGGAAAGCAAGGAAAGT 149
Db 335 AGTAGGCAGTCTGTCTGCTCTCAAAAAGGGAAGGGAAGGAGGAGGAGGGT 276
Qy 150 CGTTATTCTCCAAAGGGAATACGCTTAAATCCCTAACCCCTCGAGAGGATAAGTTGCCGA 209
Db 275 CATGAGTCATGCACCAATCATCAATTAATTAATCATGTACAGATATTTGCAAAA 216
Qy 210 CGATGTTATGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAG 256
Db 215 AAAAAATTAAAGCATAAGGAATGAAAGAGGATTTAGGTACAAAAG 169

RESULT 5
AG318648/c

LOCUS
DEFINITION
AG318648 714 bp DNA linear GSS 02-JUN-2004
Mus musculus molossinus DNA, clone:MSMg01-103007.T7, genomic survey
sequence.

ACCESSION
AG318648
VERSION
AG318648.1 GI:47891605
KEYWORDS
GSS.
SOURCE
Mus musculus molossinus
ORGANISM
Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
BAC end Sequences of Library MSMg01
Unpublished
JOURNAL
2 (bases 1 to 714)
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
Direct Submission
TITLE
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9111, Fax: 81-45-503-9170
[E-mail: hattori@gsc.riken.jp, URL: http://hgp.gsc.riken.go.jp/,
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp

PRIMERS
Sequencing : T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
source
1..714
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-103007.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match 8.1%; Score 38; DB 9; Length 714;
Best Local Similarity 51.8%; Pred. No. 4;
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 218 TGGGGGATATGAAGATAAGGAGCAGTTTACGAAGAGAGGGGTATTGGTTCTTTTCTT 277
Db 388 TGGGAGTGGATGGGTATTGAGCAGGGGAGGGGAGGATAGGGGGGGTTCCTAA 329
Qy 278 GGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCTCAAAATGAAC 337
Db 328 GGGTAAACACAGGAAGGGGATAACATTTAAATGTAATAACGAAATATCTAATAAAAA 269
Qy 338 CTGTGCAATTAAGATATGGAGAGATTTGTGCAGATCGCATTTAT 383
Db 268 ATGTAGATTTTCAGATATGGGAAATTAGATAGTGAATTTCTTTGT 223

RESULT 6
CNS0106X
LOCUS
DEFINITION
CNS0106X
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL098595
VERSION
AL098595.1 GI:5610206
KEYWORDS
GSS.

SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN03K20"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : T7"
ORIGIN
Query Match 8.0%; Score 37.8; DB 9; Length 1101;
Best Local Similarity 15.2%; Pred. No. 6.1; Mismatches 128; Indels 0; Gaps 0;
Matches 50; Conservative 151
QY 135 AAAGAAGGAAAGTCGTTATTCCTCCAAAGGAATACGCTTAATCTTAACCTCGAGAG 194
Db 749 AKADARDKRRWDKDKRRAKAAKAAADADADAKADAKADAKADADDDGRGDG 808
QY 195 GATAAGTTGCCGACGATGTTATGGGGATATGAAGATAAGGAGCAGTTTACGACAG 254
Db 809 KKKRAKDRKKKKKKKAWGDKKKAAKAAKADADAAGDAKAKRRRAGDKDKAKADAKA 868
QY 255 AGGGGTATTGTTCTTTGCTTGGGTGACCCAGGATGGAACTTAACACTAAT 314
Db 869 AKKAKADDDDAKAAATKAAKATKAKKAKKAKKKKKDKDKAKAKADKDKDD 928
QY 315 GCTCTACAATGCTCAAACTGAACCTGTGCAATTAAGATATGAGAGAGATTTGTGCAGAT 374
Db 929 KDKKDAKADKKKKKDKRAKDKKDDDKKDKAKDKDKDKADAKAKADADAKAK 988
QY 375 CGCATTTTAAAGCTAGAGGTCCGGCAAGAAACCTTTACAGAGGAACTATCAGGGGAG 434
Db 989 AKAKAKAADDAAADAKAAADAKAKAKADDAKAAKADAKADADADAKAKAAKRAK 1048
QY 435 CACNAGTTAGCGTTTTCANAGGAAAGA 463
Db 1049 AKADADAAKAKADKDKDKADKDDADKA 1077
RESULT 7
CW017266/c 604 bp DNA linear GSS 23-SEP-2004
LOCUS ZMMLC0010B24.r ZMMLC Zea mays genomic clone ZMMLC0010B24 3',
DEFINITION genomic survey sequence.
ACCESSION CW017266
VERSION CW017266.1 GI:52607659
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE 1 (bases 1 to 604)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 604)

AUTHORS Wing, R., Luo, M., Soderlund, C. and Haller, K.
TITLE ZMMLC sequences
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
Plate: 0010 row: B column: 24
Class: BAC ends.
FEATURES
source
1..604
Location/Qualifiers
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMLC0010B24"
/tissue type="immature ears"
/lab host="DH10B T1 phage resistant"
/clone_lib="ZMMLC"
/note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN
Query Match 8.0%; Score 37.6; DB 9; Length 604;
Best Local Similarity 47.1%; Pred. No. 6;
Matches 115; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 118 AAATTAATCGACGTGGAAAGGAAGAAAGTCGTTATTCCTCCAAAGGAATACGCCTTA 177
Db 479 ATAATACCAAGTTGTAAACTATAGAGAAGGGGATACTCTCTCATGTACTTTGTCTCC 420
QY 178 ATCCTAACCTCGAGAGGATAAGTTGCCGACGATGTTATGGGGATATGAAGATAAGG 237
Db 419 ATTTATTATCATGTGAATTAACCAACCCAAAAGTGGCTAATTACGAATGAGACATGA 360
QY 238 AGCAGTTTAGCAAGAAGAGGGGTTATTGGTTCTTTTGGTTGGTTCACCCAGGATGGGAT 297
Db 359 ACNACTTTGGCATCTGCTGGAGTTTATATGATTTGTCATCTCCTAAAGATAGTGTAT 309
QY 298 GGAAGACTTAACACTAATGCTCTACAATGCCTCAATGAACCTGTGCAATTAGATATGA 357
Db 299 AAAAGAAATATACAAATGCTGAATAGGCATTAACTTAATTTTGGAAAATTAAGAATTGA 240
QY 358 GAGA 361
Db 239 GAAA 236
RESULT 8
AG04615/c 808 bp DNA linear GSS 03-JUN-2004
LOCUS Mus musculus molossinus DNA, clone:MSMg01-262116.T7, genomic survey
DEFINITION sequence.
ACCESSION AG04615
VERSION AG04615.1 GI:48047301
KEYWORDS GSS.
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus
REFERENCE 1
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
BAC end Sequences of Library MSMg01
2 (bases 1 to 808)
Direct Submission
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: hattori@gsc.riken.jp, URL: http://hgp.gsc.riken.go.jp/.

collaboration with the European Drosophila Genome Project (BDGP) - <http://www.edgp.ebi.ac.uk>.. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CSHP (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES

Location/Qualifiers
1..1019
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN03D10"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : SP6"

ORIGIN

```
Query Match          7.9%; Score 37; DB 9; Length 1019;
Best Local Similarity 41.8%; Pred. No. 10; Mismatches 125; Indels 0; Gaps 0;
Matches 97; Conservative 10;

QY 233 TAAGGAGCAGTTAGCAAGAGAGGGTTATTGGTTCTTTTGGGTGACCCAGGAT 292
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 623 TAAATWTAATAAACAAGATAGMGGGTTTAAAGGAATTTATGGGAATTTTAAAGAAGAT 564
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 293 GGGATGGAACTTAACATAGCTCTACATGCCTCAATGAACTGTCGAAATTAAGAT 352
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 563 AGAAATGAAATGGAAATATAAAGAGAGGAAATAATAAAGATATAAAGAAATTAAGAW 504
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 353 ATGGAGAGAGATTTGCGAGATCGCATTTATAGGCTAGAGGTCGCGCAAGAAACCTT 412
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 503 AWATGAAATATGAAGAAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 444
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 413 ACAGAGGAACTATCAGGGGAGCACAAGGTAGCGTTTTTCAAGAGAGAAAGAA 464
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 443 ANAAGAAWAAATAAAGAAGAAAGAAAGAAATGAGAAAGAAAGAAAGTGAAGGA 392
```

RESULT 11

CNS005TE/c

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence TET3 end of BAC #

BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL060767

AL060767.1 GI:4943573

GSS.

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 997)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammeter in Piter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

p1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers
1..997
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR12K22"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

```
Query Match          7.8%; Score 36.8; DB 9; Length 997;
Best Local Similarity 17.2%; Pred. No. 12;
Matches 78; Conservative 154; Mismatches 222; Indels 0; Gaps 0;

QY 12 AGACTGGAAATCAGAAAGAAATCTTATAGACCATTTTCTGAGGAATCGCTCAACC 71
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 931 AAARRRARRRAGRRRAGRRSAGGRRGRRRAAARARARARARARARARARARAR 872
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 72 AGCAGGTTATCACCTCAGAGTGGCGAGAGCGCTTTTCTTAAGGGGAAATTAATCGCGT 131
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 871 AKAARAGARRRRRRRRRRRRRAGRRRAGGRRGRGRRRRARRRAGARARARAR 812
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 132 GGAAGAGGAAGAAAGTCGTTATTCCTCCAGGGAATACGCTTAATCCTAACCTCGA 191
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 811 GRRRAARARARRRRRRRRRRRARGRRRRRGGRRRRRRRRRRRARRARARARA 752
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 192 GAGGATAAGTTCGCGAGCATGTTATGGGGATATGAAGTAAGGAGCAGTTTAGCAAG 251
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 751 GARAAARRRRARRRAGAAARRRRRRRRRRRRRRRAGARRRRRRRRRAGAGARRRR 692
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 252 AGAAGGGGTTATTGGTTCTTTTGGTTTGACCCAGGATGGGATGGAACCTTAACACT 311
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 691 RRGAGARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 632
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 312 AATGCTCTACAATGCCTCAATGAACCTGTGCAATTAAGATATGAGAGAGATTTGTGCA 371
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 631 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 572
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 372 GATCGCTTTATAGGCTAGAGGTCGCGCAAGAACCTTACAGAGGAACCTATCAGGG 431
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 571 RGRARRRARRRARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 512
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 432 GAGCAACAGGTTAGCGTTTTCAGAGAGAAAGAAA 465
    ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 511 RGGARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 478
```

RESULT 12

CE439533

LOCUS

DEFINITION

tigr-gss-dog-17000335920055 Dog Library Canis familiaris genomic,

genomic survey sequence.

CE439533

CE439533.1 GI:36720347

GSS.

Canis familiaris (dog)

Canis familiaris

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

1 (bases 1 to 409)

Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,

Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and

Venter,J.C.

The dog genome: survey sequencing and comparative analysis

Science 301 (5641), 1898-1903 (2003)

22875432

14512627

CONTACT: Kirkness EF

The Institute for Genomic Research

Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,

Rockville, MD 20850, USA

Tel: 301-838-0200

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 07:12:26 ; Search time 484 Seconds
(without alignments)
6103.716 Million cell updates/sec

Title: US-08-957-709A-70
Perfect score: 471
Sequence: 1 agctacttccagactggaa.....caagaagaagaactctag 471

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.2	8.5	2964	20	US-10-784-986-3
2	38.6	8.2	89047	18	US-10-672-787-34
3	37.2	7.9	110021	21	US-10-461-862-83
4	36.2	7.7	644	13	US-10-027-632-224178
5	36.2	7.7	644	17	US-10-027-632-224178
6	35	7.4	3673778	16	US-10-312-841-1
7	34.8	7.4	570	9	US-09-864-761-7422

8	34.8	7.4	570	9	US-09-864-761-8565	Sequence 8565, Ap
9	34.6	7.3	649	13	US-10-027-632-226698	Sequence 226698,
10	34.6	7.3	649	17	US-10-027-632-226698	Sequence 226698,
11	34.6	7.3	90435	19	US-10-322-696-160	Sequence 160, App
12	34.2	7.3	449	14	US-10-123-155-224	Sequence 224, App
13	34.2	7.3	449	15	US-10-146-731-224	Sequence 224, App
14	34.2	7.3	449	15	US-10-140-472-224	Sequence 224, App
15	34.2	7.3	449	15	US-10-141-761-224	Sequence 224, App
16	34.2	7.3	449	16	US-10-142-885-224	Sequence 224, App
17	34.2	7.3	449	16	US-10-158-790-224	Sequence 224, App
18	34.2	7.3	449	17	US-10-137-871-224	Sequence 224, App
19	34.2	7.3	449	17	US-10-140-923-224	Sequence 224, App
20	34.2	7.3	449	17	US-10-141-756-224	Sequence 224, App
21	34.2	7.3	449	17	US-10-141-759-224	Sequence 224, App
22	34.2	7.3	449	17	US-10-140-805-224	Sequence 224, App
23	34.2	7.3	449	17	US-10-140-864-224	Sequence 224, App
24	34.2	7.3	449	18	US-10-142-426-224	Sequence 224, App
25	34	7.2	1691139	14	US-10-067-514-1	Sequence 1, Appli
26	34	7.2	1691139	17	US-10-419-723-1	Sequence 1, Appli
27	33.8	7.2	2256646	19	US-10-470-565-1	Sequence 1, Appli
28	33.6	7.1	405	10	US-09-975-719-242	Sequence 242, App
29	33.6	7.1	765	10	US-09-975-719-240	Sequence 240, App
30	33.6	7.1	12269	15	US-10-311-455-1384	Sequence 1384, Ap
31	33.6	7.1	12269	18	US-10-221-714A-194	Sequence 194, App
32	33.6	7.1	38844	13	US-10-060-333-3	Sequence 3, Appli
33	33.6	7.1	42235	10	US-09-975-719-1	Sequence 1, Appli
34	33.4	7.1	700	17	US-10-398-221-2976	Sequence 2976, Ap
35	33.4	7.1	746	20	US-10-425-115-20392	Sequence 20392, A
36	33.4	7.1	2107	20	US-10-739-930-4920	Sequence 4920, Ap
37	33.4	7.1	114931	13	US-10-087-192-862	Sequence 862, App
38	33.2	7.0	985	18	US-10-425-114-6662	Sequence 6662, Ap
39	33.2	7.0	1945	20	US-10-425-115-101174	Sequence 101174,
40	33	7.0	326	11	US-09-732-627A-4547	Sequence 4547, Ap
41	33	7.0	2731748	19	US-10-297-465A-1	Sequence 1, Appli
42	32.8	7.0	544	14	US-10-184-644-274	Sequence 274, App
43	32.8	7.0	544	14	US-10-184-634-274	Sequence 274, App
44	32.8	7.0	775	14	US-10-123-155-120	Sequence 120, App
45	32.8	7.0	775	15	US-10-146-731-120	Sequence 120, App

ALIGNMENTS

RESULT 1
US-10-784-986-3
; Sequence 3, Application US/10784986
; Publication No. US20040229311A1
; GENERAL INFORMATION:
; APPLICANT: HIRANO, Seiko
; APPLICANT: YASUEDA, Hisaehi
; TITLE OF INVENTION: Novel lysine decarboxylase gene and method for
; TITLE OF INVENTION: producing L-lysine
; FILE REFERENCE: US-109
; CURRENT APPLICATION NUMBER: US/10/784,986
; PRIOR FILING DATE: 2004-02-25
; PRIOR APPLICATION NUMBER: JP 2003-47185
; PRIOR FILING DATE: 2003-02-25
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2964
; TYPE: DNA
; ORGANISM: Methylophilus methylotrophus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (684)..(2930)
US-10-784-986-3

Query Match 8.5%; Score 40.2; DB 20; Length 2964;
Best Local Similarity 46.3%; Pred. No. 0.091;
Matches 132; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
QY 87 CAGAGTGGCAGAGAGGCTTTTGTAAAGGGGAAATTAATCGACGTGGAAAGGAAA 146

```
Db 195 CATAGTTACCCCAAGCAATTGACCCGACGTCGTTTTCGAGGTCTCCGCAAGGCTA 254
Qy 147 AGTCGTTATCTTCAAGGGAATACGCTTAATCTTAACCCCTCGAGAGGATAAAGTTGCC 206
Db 255 TTGCGTATTCCTCCCTAACTCATTTGCACTGGCGGCACGGTAGAGATTTTCCGTTATTC 314
Qy 207 CGACATCTATGGGGGATATCAAGATAAGGAGCAGTTTAGCAAGAGAAGGGTTATTGG 266
Db 315 TCGCTCTGTACTGACTGTATGCTCGGCAAGTCGACTTATGCGCGTTGCGGCATTATCGT 374
Qy 267 TTCTTTTGTCTGGGTTGACCCAGGATGGATGGAAACTTAACACTAATGCTCTACAATGC 326
Db 375 CAACGTCACCCCTTTGACCAAGAGTGGGAGGCTATGTCACTAGAGTTTCAGCAACAC 434
Qy 327 CTCAAAATGAACCTGTCGAATTAAAGATATGGAGAGAGATTGTGCA 371
Db 435 CACACCGCTACCCGCAAAATTTATGCTGGCGAAGGCTGTGGCA 479

RESULT 2
US-10-672-787-34
; Sequence 34, Application US/10672787
; Publication No. US20040067554A1
; GENERAL INFORMATION:
; APPLICANT: LAGACE, Robert, E.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BERG, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: ELITRA.025C1
; CURRENT APPLICATION NUMBER: US/10/672,787
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 09/596,002
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 34
; LENGTH: 89047
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 66874
; OTHER INFORMATION: a or g or c or t, unknown, or other
US-10-672-787-34

Query Match 8.2%; Score 38.6; DB 18; Length 89047;
Best Local Similarity 47.3%; Pred. No. 1.5;
Matches 116; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 152 TTATTCTCTCAAGGGAATACGCTTAATCTTAACCCCTCGAGAGGATAAAGTTGCCGAGC 211
Db 57078 TCATTCCACAAATTCATTGCTTTGGCGGCACACTGTTGAGTATTTTCGCATCCCAAGAG 57137
Qy 212 ATGTTATGGGGATATGAAGATAAGGAGCAGTTTAGCAGAGAAGGGTTATTGTTCTT 271
Db 57138 ATGTACTGACGACTCTGTTTGGGCAATCCACTTATGACGCTGTGGAATCATGTCAATG 57197
Qy 272 TTGCTTGGGTTGACCCAGGATGGATGGAAACTTAACACTAATGCTCTACAATGCTCAA 331
Db 57198 TCAGGCCACTTGAGCTGTAATGGGAGGGCATGTAACCTTAGAATTTAGCAATACCACCA 57257
Qy 332 ATGAACCTGTCGAATTAAGATATGGAGAGAGATTTGTGCAGATCGCATTTTATAAGCTTAC 391
Db 57258 ATTTGCGACGACGCTATCTATGCTGTGAGGGTGTGGCAGACAGATGCTATTTTCCCAATCAG 57317
Qy 392 AGGCT 396
Db 57318 ATGCT 57322

RESULT 3
US-10-461-862-83/c
```

```
; Sequence 83, Application US/10461862
; Publication No. US20050090434A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001800
; CURRENT APPLICATION NUMBER: US/10/461,862
; CURRENT FILING DATE: 2003-06-13
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83
; LENGTH: 110021
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(110021)
; OTHER INFORMATION: n = A,T,C or G
US-10-461-862-83

Query Match 7.9%; Score 37.2; DB 21; Length 110021;
Best Local Similarity 52.6%; Pred. No. 4.7;
Matches 81; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 143 GAAAAGTCGTTATTCTCTCAAGGGAATACGCTTAAATCCTTAACCCCTCGAGAGGATAAAGT 202
Db 81964 GGACATTCATCATGATACCACTTAGAATTTCTTCCACCTGACCATGAGATGATGAATA 81905
Qy 203 TGCCCGCAGATGTTATGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAAGGGGTTA 262
Db 81904 GGCCACACCGTTATGTTTTCTTAAAAATGAAAAGGCGAGTTGAGAAAACAGAACAGTCA 81845
Qy 263 TTGGTCTTTTGGTGGTTGACCCAGCATGGGA 296
Db 81844 TTCTGTCCTTTTTCAGTGAGGAAGCTGAGGTAGGTA 81811

RESULT 4
US-10-027-632-224178
; Sequence 224178, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 224178
; LENGTH: 644
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-224178

Query Match 7.7%; Score 36.2; DB 13; Length 644;
Best Local Similarity 53.1%; Pred. No. 0.86;
```

Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 4 CTAATCCAGCTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGATCG 63
Db 432 CTACAACTAATCGAGATATTAAGAGGAGGAAAGGATTAATCCCATCTTTGGTAGAAGG 491
QY 64 CTCACACCGAGCGTTATGACCTCAGAGTGGGAGAGAGGCTTTTGTAAAGGGGAATTA 123
Db 492 CTTCAAGCAGTACAGATAAGCCCTTAAGGTGCTTAGCATCCTTTTAGATAAGGAATATA 551
QY 124 ATCGACGTGGAAAGGAGGAAAG 148
Db 552 ATAGATGCCAGATAAGAAAG 576

RESULT 5
US-10-027-632-224178
; Sequence 224178, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 224178
; LENGTH: 644
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-224178

Query Match 7.7%; Score 36.2; DB 17; Length 644;
Best Local Similarity 53.1%; Pred. No. 0.86;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 4 CTAATCCAGCTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGATCG 63
Db 432 CTACAACTAATCGAGATATTAAGAGGAGGAAAGGATTAATCCCATCTTTGGTAGAAGG 491
QY 64 CTCACACCGAGCGTTATGACCTCAGAGTGGGAGAGAGGCTTTTGTAAAGGGGAATTA 123
Db 492 CTTCAAGCAGTACAGATAAGCCCTTAAGGTGCTTAGCATCCTTTTAGATAAGGAATATA 551
QY 124 ATCGACGTGGAAAGGAGGAAAG 148
Db 552 ATAGATGCCAGATAAGAAAG 576

RESULT 6
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO

; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 7.4%; Score 35; DB 16; Length 3673778;
Best Local Similarity 59.6%; Pred. No. 1.2e+02;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 208 GACGATGTTATGGGGATATGAGATAAGGACAGTTCACGAGAGAGGTTTACGAGAGAGAGGGTTATTGGT 267
Db 3478762 GAAGTGGTTTGGGAAATTTGAAGATATTGGGATATTTAGAAGGTTTAAGGGGATTTAGTT 3478821
QY 268 TCTTTTGTGGTTCACCCAGGATGGGATGGAAACTTA 306
Db 3478822 TATTTGTGGTAAAGTGTGGGAGTGAAGTAGGTAA 3478860

RESULT 7
US-09-864-761-7422
; Sequence 7422, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21

;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 7422
;; LENGTH: 570
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC018367.3
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
US-09-864-761-7422

Query Match 7.4%; Score 34.8; DB 9; Length 570;
Best Local Similarity 47.0%; Pred. No. 2.3; Mismatches 0; Gaps 0;
Matches 108; Conservative 0; Indels 122; Indels 0; Gaps 0;
QY 155 TTCTCCAAAGGAATACGCTTAATCCTAAACCTCGAGAGGATAAAGTTGCCCGACGATG 214
Db |||||
QY 229 TGCCAGGAGAGACTCAACAACATCCCATCAACTGGATGTTAAGATTGCCACCTGGAC 288
Db |||||
QY 215 TTATGGGGGATATGAAGATAAGAGCAGTTTATAGCAAGAGAGGGGTTATGTTCTTTTG 274
Db |||||
QY 289 ACTTGTGCTCTCTTCTTACATTTAAGTCAACAGGCTAAAAAGGAGTTACAGTGTGGCTA 348
QY 275 CTGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTTACATGCTCAATG 334
Db |||||
QY 349 AGGTGATTGACCCAGACTATCATGATGAAATCAGCTACTCTCCACATGAAGTAAAGG 408
QY 335 AACCTGTGCAATTAAGATATGAGAGAGAGATTTGTGCGAGATCCCATTTATA 384
Db |||||
QY 409 AAGAGTACGCATGGAATACAGAGATCCATTAGGGCATCTCTTAGTATTA 458

RESULT 8
US-09-864-761-8565
;; Sequence 8565, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wenheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aomic-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 8565
;; LENGTH: 570
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL157957.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
US-09-864-761-8565
Query Match 7.4%; Score 34.8; DB 9; Length 570;
Best Local Similarity 47.0%; Pred. No. 2.3; Mismatches 0; Gaps 0;
Matches 108; Conservative 0; Indels 122; Indels 0; Gaps 0;
QY 155 TTCTCCAAAGGAATACGCTTAATCCTAAACCTCGAGAGGATAAAGTTGCCCGACGATG 214
Db |||||
QY 229 TGCCAGGAGAGACTCAACAACATCCCATCAACTGGATGTTAAGATTGCCACCTGGAC 288
Db |||||
QY 215 TTATGGGGGATATGAAGATAAGAGCAGTTTATAGCAAGAGAGGGGTTATGTTCTTTTG 274
Db |||||
QY 289 ACTTGTGCTCTCTTCTTACATTTAAGTCAACAGGCTAAAAAGGAGTTACAGTGTGGCTA 348
QY 275 CTGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTTACATGCTCAATG 334
Db |||||
QY 349 AGGTGATTGACCCAGACTATCATGATGAAATCAGCTACTCTCCACATGAAGTAAAGG 408
QY 335 AACCTGTGCAATTAAGATATGAGAGAGAGATTTGTGCGAGATCCCATTTATA 384
Db |||||
QY 409 AAGAGTACGCATGGAATACAGAGATCCATTAGGGCATCTCTTAGTATTA 458

RESULT 9
US-10-027-632-226698
;; Sequence 226698, Application US/10027632
;; Publication No. US20020198371A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, David G.
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; TITLE OF INVENTION: Polymorphisms in the Human Genome
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 226698
LENGTH: 649
TYPE: DNA
ORGANISM: Human
US-10-027-632-226698

Query Match
Best Local Similarity 7.3%; Score 34.6; DB 13; Length 649;
Matches 97; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 83 ACCTCAGAGTGGCCAGAGAGCGCTTTTAAAGGGGAAATTAATCGACGTCGGAAGGAAG 142
DB 189 ATCTGAGTGGCTTATGGAAGAATTTGGTAAGAGGAGGAAAGAGTAAGGGAATAATGGA 248
QY 143 GAAAGTCGTTATTCCTCCAGGGAATACGCCCTTAATCCTTAACCTCGAGAGGATAAAGT 202
DB 249 TGAAGCTGCGTATCAGAGAAYGTGAGAGGACAAAGACCAAGACACAGGTGGAAGATT 308
QY 203 TGCCCGACGATGTTATGCGGGGATATGAAGATAAGGAGCAGTTTACGAAGAGAGGGGTTA 262
DB 309 TGCCCGAGTTGGTAACAGAGGAGGAAATGGAATGATGACTGTGGGAGTAGATGTGTTG 368
QY 263 TTGGTTCTTTTGGCTGGGTG 283
DB 369 TCATTTTGATGGCTGGACATG 389

RESULT 10
US-10-027-632-226698
Sequence 226698, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMERASE: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 226698
LENGTH: 649
TYPE: DNA
ORGANISM: Human
US-10-027-632-226698

Query Match
Best Local Similarity 7.3%; Score 34.6; DB 17; Length 649;
Matches 97; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 83 ACCTCAGAGTGGCCAGAGAGCGCTTTTAAAGGGGAAATTAATCGACGTCGGAAGGAAG 142

DB 189 ATCTGAGTGGCTTATGGAAGAATTTGGTAAGAGGAGGAAAGAGTAAGGGAATAATGGA 248
QY 143 GAAAGTCGTTATTCCTCCAGGGAATACGCCCTTAATCCTTAACCTCGAGAGGATAAAGT 202
DB 249 TGAAGCTGCGTATCAGAGAAYGTGAGAGGACAAAGACCAAGACACAGGTGGAAGATT 308
QY 203 TGCCCGACGATGTTATGCGGGGATATGAAGATAAGGAGCAGTTTACGAAGAGAGGGGTTA 262
DB 309 TGCCCGAGTTGGTAACAGAGGAGGAAATGGAATGATGACTGTGGGAGTAGATGTGTTG 368
QY 263 TTGGTTCTTTTGGCTGGGTG 283
DB 369 TCATTTTGATGGCTGGACATG 389

RESULT 11
US-10-322-696-160/c
Sequence 160, Application US/10322696
Publication No. US20040166490A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
TITLE OF INVENTION: NOVEL THERAPEUTIC TARGETS IN CANCER
FILE REFERENCE: 529452001200
CURRENT APPLICATION NUMBER: US/10/322,696
CURRENT FILING DATE: 2003-10-17
NUMBER OF SEQ ID NOS: 186
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 160
LENGTH: 90435
TYPE: DNA
ORGANISM: Homo sapiens
US-10-322-696-160

Query Match
Best Local Similarity 7.3%; Score 34.6; DB 19; Length 90435;
Matches 55; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 379 TTTATAAGGCTGAGGCGTCGGCAAGAAACCCCTTACAGAGAGAACTATCAGGGAGCACA 438
DB 81318 TTTCAATGCTAAAGAGAGAGCAAAACCAATTAATAGAAAAAATCAATGAAGCTAA 81259
QY 439 AGGTTAGCGTTTCAAGAGAGAAAGAACT 467
DB 81258 AGCTTAGTCTTTTGAAGAGAGTAATACAAT 81230

RESULT 12
US-10-123-155-224
Sequence 224, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
FILE REFERENCE: P33301C30
CURRENT APPLICATION NUMBER: US/10/123,155

Db 202 PRATLATLATRAQTAVATTANTSSPMSTRPSKHPMSDTPAASFPVPPMRPQAQGPISQVS 261

QY 303 CTTAACACT 311

Db 262 VDQPVVNTT 270

RESULT 15

US-10-141-761-224
; Sequence 224, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 224
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-141-761-224

Query Match 7.3%; Score 34.2; DB 15; Length 449;
Best Local Similarity 20.5%; Pred. No. 3.2;
Matches 51; Conservative 65; Mismatches 133; Indels 0; Gaps 0;
QY 63 GCTCCACACGAGGTTATGACCTCAGAGTGGCGCAGAGGCTTTTGTAAAGGGAAATT 122
Db 22 SNDPRNFVPKMKWGLVKRNASVETVDNKTSEDVTMAAASPVLTTRKTSAAHLNSMEVTT 81
QY 123 AATCGAGCGTGAAGAAAGTGGTTATTCTCCAAAGGAATACGCCCTTAATCCT 182
Db 82 EDTSRDVBEPATSGVAADGVTIAPTAVASSTTAASITTAASSMTVASSAPTTAASST 141
QY 183 AACCCCTCGAGAGGATAAAGTGGCCGAGATGTTATGGGGATATGAGATAGAGGACAG 242
Db 142 VASIAPTTAASSMTAASSTPMTLALPAPTSTSTGRTPSTTATGHPSLSTALAQVPKSSAL 201
QY 243 TTATGCAAGAGAGGGTTATGCTTCTTTGCTGGTGGTACCCAGGATGGGATGAAA 302
Db 202 PRATLATLATRAQTAVATTANTSSPMSTRPSKHPMSDTPAASFPVPPMRPQAQGPISQVS 261
QY 303 CTTAACACT 311
Db 262 VDQPVVNTT 270

Search completed: July 2, 2005, 09:10:49
Job time : 493 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 06:11:49 ; Search time 145 Seconds
(without alignments)
5315.078 Million cell updates/sec

Title: US-08-957-709A-70
Perfect score: 471
Sequence: 1 atgctactccagactggaa.....caagagagaagaactcttag 471

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A-COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B-COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A-COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B-COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTUS-COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfilee1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	471	100.0	471	4	US-09-399-003-70
2	316	67.1	740	3	US-08-922-774-42
3	316	67.1	740	3	US-09-632-711-42
4	316	67.1	740	3	US-09-632-703B-42
5	316	67.1	740	3	US-09-632-702-42
6	316	67.1	740	4	US-09-399-003-42
7	96.4	20.5	129	3	US-08-822-774-40
8	96.4	20.5	129	3	US-09-632-711-40
9	96.4	20.5	129	3	US-09-632-703B-40
10	96.4	20.5	129	3	US-09-632-702-40
11	96.4	20.5	129	4	US-09-399-003-40
12	82	17.4	84	3	US-08-822-774-36
13	82	17.4	84	3	US-09-632-711-36
14	82	17.4	84	3	US-09-632-703B-36
15	82	17.4	84	3	US-09-632-702-36
16	82	17.4	84	4	US-09-399-003-36
17	54.2	11.5	7218	1	US-08-232-463-14
18	38.6	8.2	579	4	US-09-540-236-1761
19	38.6	8.2	89047	4	US-09-596-002-34
20	37.2	7.9	561	4	US-09-248-796A-1970
21	37.2	7.9	846	4	US-09-248-796A-1971
22	35.2	7.5	267	4	US-09-248-796A-11979
23	34.4	7.3	601	4	US-09-949-016-204563
24	34.4	7.3	601	4	US-09-949-016-204564
25	34.4	7.3	601	4	US-09-949-016-204565
26	34.4	7.3	87734	4	US-09-949-016-17521
27	33.6	7.1	405	3	US-09-199-637A-242

28	33.6	7.1	765	3	US-09-199-637A-240	Sequence 240, App
29	33.6	7.1	3844	3	US-09-734-675-3	Sequence 3, Appl
30	33.6	7.1	42235	3	US-09-199-637A-1	Sequence 1, Appl
31	32.6	6.9	601	4	US-09-949-016-26855	Sequence 26855, A
32	32.6	6.9	601	4	US-09-949-016-55850	Sequence 55850, A
33	32.6	6.9	227390	4	US-09-949-016-12201	Sequence 12201, A
34	32.6	6.9	227391	4	US-09-949-016-13365	Sequence 13365, A
35	32.2	6.8	616	4	US-09-573-080A-417	Sequence 417, App
36	32.2	6.8	25656	4	US-09-949-016-13022	Sequence 13022, A
37	32	6.8	594	4	US-09-328-352-539	Sequence 539, App
38	32	6.8	23781	4	US-09-949-016-13446	Sequence 13446, A
39	32	6.8	39601	4	US-09-949-016-16045	Sequence 16045, A
40	32	6.8	147382	4	US-09-949-016-14624	Sequence 14624, A
41	31.6	6.7	491	3	US-08-896-164-11	Sequence 11, Appl
42	31.6	6.7	399	4	US-09-621-976-8976	Sequence 8976, App
43	31.6	6.7	609	4	US-09-252-991A-1091	Sequence 1091, App
44	31.6	6.7	3101	3	US-09-221-017B-87	Sequence 87, Appl
45	31.6	6.7	187136	4	US-09-949-016-17231	Sequence 17231, A

ALIGNMENTS

RESULT 1
US-09-399-003-70
; Sequence 70, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Hogrefe, Holly
; APPLICANT: Hansen, Connie J
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
; FILE REFERENCE: 4121.0116-02
; CURRENT APPLICATION NUMBER: US/09/399,003
; CURRENT FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/ US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
US-09-399-003-70

Query Match 100.0%; Score 471; DB 4; Length 471;
Best Local Similarity 100.0%; Pred. No. 2.7e-149;
Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGCTACTTCCAGCTGGAAATACAGAAAGAAATACCTATAGAGCAATTTCTGAGAA	60
Db	1	ATGCTACTTCCAGCTGGAAATACAGAAAGAAATACCTATAGAGCAATTTCTGAGAA	60
QY	61	TCGCTCCACACAGCAGGTTATGACCTCAGAGTGGGAGAGAGGCTTTGTTAAGGGGAAA	120
Db	61	TCGCTCCACACAGCAGGTTATGACCTCAGAGTGGGAGAGAGGCTTTGTTAAGGGGAAA	120
QY	121	TTAATCGAGCTGGAAAGGAAAGTGGTTATTTCTTCAAGGGAATACGCTTAAATC	180
Db	121	TTAATCGAGCTGGAAAGGAAAGTGGTTATTTCTTCAAGGGAATACGCTTAAATC	180
QY	181	CTAACCTCGAGAGGATAAGTTCCCGAGCATGTTATGGGGATATGAGATAGAGGC	240
Db	181	CTAACCTCGAGAGGATAAGTTCCCGAGCATGTTATGGGGATATGAGATAGAGGC	240
QY	241	AGTTTACGAGAGAGAGGTTATTGGTTCTTTTGGTTGACCCAGGATGGGATGGA	300
Db	241	AGTTTACGAGAGAGAGGTTATTGGTTCTTTTGGTTGACCCAGGATGGGATGGA	300

301	AACTTAACACTAATGCTCTTCAACAATGCCCTCAAACTGAACCTGTGCGAATTTAGATATGGAGAG	360
Qy		
301	AACTTTAACACTAATGCTCTTCAAAATGCCCTCAAACTGAACCTGTGCGAATTTAGATATGGAGAG	360
Db		
361	AGATTGTGCAGATCGCATTTTATAAGGCTAGAGGGTCCGGCAGAAACCCCTTACAGAGGA	420
Qy		
361	AGATTGTGCAGATCGCATTTTATAAGGCTAGAGGGTCCGGCAGAAACCCCTTACAGAGGA	420
Db		
421	AACTATCAGGGGAGCACAAAGTGTACGTTTTTCAAAGAGAAAGAAACCTCTAG	471
Qy		
421	AACTATCAGGGGAGCACAAAGTGTACGTTTTTCAAAGAGAAAGAAACCTCTAG	471
Db		

```

RESULT 2
US-08-822-774-42
; Sequence 42, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
US-08-822-774-42

```

```

336 ACCTGTCGAATTAAGATATCGAGAGAGATTTGTGCAGATCCGATTTTATAAGGCTAGAGGG 395
181 ACCTGTCGAATTAAGATATCGAGAGAGATTTGTGCAGATCCGATTTTATAAGGCTAGAGGG 240
396 TCCGGCAAGAAACCCCTTACAGAGAAACTATCAGGGGAGCAACAAGTTTAGCGTTTTCAAA 455
241 TCCGGCAAGAAACCCCTTACAGAGAAACTATCAGGGGAGCAACAAGTTTAGCGTTTTCAAA 300
456 GAGAAAGAAACTCTAG 471
301 GAGAAAGAAACTCTAG 316

RESULT 3
US-09-632-711-42
; Sequence 42, Application US/09632711
; Patent No. 6333165
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymase Enhancing Factor (PEP)
; Extracts, PEP Protein Complexes, Isolated PEP Proteins,
; and Methods for Purifying and Identifying Same
;
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,711
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/922,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-632-711-42

```

QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACATAATGCTCTACAATGCGCTCAAAATGA 335
Db 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACATAATGCTCTACAATGCGCTCAAAATGA 180
QY 336 ACCTGTGCGAATTAAGATATGAGAGAGATTTGTGCGAGATCGCATTTTATAAGGCTAGAGGG 395
Db 181 ACCTGTGCGAATTAAGATATGAGAGAGATTTGTGCGAGATCGCATTTTATAAGGCTAGAGGG 240
QY 396 TCCGGCAAGAACCCCTTACAGAGAACTATACAGGGAGCACAAGGTAGCGTTTTCAAA 455
Db 241 TCCGGCAAGAACCCCTTACAGAGAACTATACAGGGAGCACAAGGTAGCGTTTTCAAA 300
QY 456 GAGAAAGAACTCTAG 471
Db 301 GAGAAAGAACTCTAG 316
RESULT 4
US-09-632-703B-42
; Sequence 42, Application US/09632703B
; Patent No. 6379553
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF
; Protein Complexes, Isolated PEF Proteins, and Methods for P
; Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,703B
; FILING DATE: 24-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BARKER, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 4121.0116-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 849-6613
; TELEFAX: (650) 849-6666
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-632-703B-42
Query Match 67.1%; Score 316; DB 3; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.2e-96;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCCAAGGGAATAGCGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCGACCATGT 215
Db 1 TCCTCCAAGGGAATAGCGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCGACCATGT 60
QY 216 TATGGGGGATATGAAGATAAGGAGCAGATTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 275

Db 61 TATGGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGGGGTTATTGGTTCTTTTGC 120
QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACATAATGCTCTACAATGCGCTCAAAATGA 335
Db 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACATAATGCTCTACAATGCGCTCAAAATGA 180
QY 336 ACCTGTGCGAATTAAGATATGAGAGAGATTTGTGCGAGATCGCATTTTATAAGGCTAGAGGG 395
Db 181 ACCTGTGCGAATTAAGATATGAGAGAGATTTGTGCGAGATCGCATTTTATAAGGCTAGAGGG 240
QY 396 TCCGGCAAGAACCCCTTACAGAGAACTATACAGGGAGCACAAGGTAGCGTTTTCAAA 455
Db 241 TCCGGCAAGAACCCCTTACAGAGAACTATACAGGGAGCACAAGGTAGCGTTTTCAAA 300
QY 456 GAGAAAGAACTCTAG 471
Db 301 GAGAAAGAACTCTAG 316
RESULT 5
US-09-632-702-42
; Sequence 42, Application US/09632702
; Patent No. 6444428
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; Lenahan, P.L.L.C.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,702
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-632-702-42
Query Match 67.1%; Score 316; DB 3; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.2e-96;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCCAAGGGAATAGCGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCGACCATGT 215

Db 1 TCCTCCAAGGAATACGCTTAATCTAAACCTCGAGAGATAAAGTTGCCCGACGATGT 60
QY 216 TATGGGGGATATCAAGATTAAGAGCAGTTTAGCAAGAGAGGGTTATTGGTCTTTTGC 275
Db 61 TATGGGGGATATGAAGATTAAGAGCAGTTTAGCAAGAGAGGGTTATTGGTCTTTTGC 120
QY 276 TTGGGTTGACCCAGGATGGATGGAACCTTAACACTAAATGCTCTACAATGCCTCAATGA 335
Db 121 TTGGGTTGACCCAGGATGGATGGAACCTTAACACTAAATGCTCTACAATGCCTCAATGA 180
QY 336 ACCTGTGCAATTAAGATATGGAGAGAGATTGTGTCAGATCGCATTTTAAAGGCTAGAGGG 395
Db 181 ACCTGTGCAATTAAGATATGGAGAGAGATTGTGTCAGATCGCATTTTAAAGGCTAGAGGG 240
QY 396 TCCGGCAAGAAACCTTTACAGAGGAAACTATCAGGGGAGCACAAGTTAGCGTTTTCAAA 455
Db 241 TCCGGCAAGAAACCTTTACAGAGGAAACTATCAGGGGAGCACAAGTTAGCGTTTTCAAA 300
QY 456 GAGAAAGAACTCTAG 471
Db 301 GAGAAAGAACTCTAG 316

RESULT 6
US-09-399-003-42
; Sequence 42, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Hogrefe, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
; TITLE OF INVENTION: Isolated PEF Proteins, and Methods for Purifying and Identifying
; FILE REFERENCE: 4121.0116-02
; CURRENT APPLICATION NUMBER: US/09/399,003
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/ US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(740)
; OTHER INFORMATION: "n" is a, t, g, or c
US-09-399-003-42

Query Match 67.1%; Score 316; DB 4; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.2e-96;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCCAAGGGAATACGCTTAATCTAAACCTCGAGAGATAAAGTTGCCCGACGATGT 215
Db 1 TCCTCCAAGGGAATACGCTTAATCTAAACCTCGAGAGATAAAGTTGCCCGACGATGT 60
QY 216 TATGGGGGATATGAAGATTAAGAGCAGTTTAGCAAGAGAGGGTTATTGGTCTTTTGC 275
Db 61 TATGGGGGATATGAAGATTAAGAGCAGTTTAGCAAGAGAGGGTTATTGGTCTTTTGC 120
QY 276 TTGGGTTGACCCAGGATGGATGGAACCTTAACACTAAATGCTCTACAATGCCTCAATGA 335
Db 121 TTGGGTTGACCCAGGATGGATGGAACCTTAACACTAAATGCTCTACAATGCCTCAATGA 180
QY 336 ACCTGTGCAATTAAGATATGGAGAGAGATTGTGTCAGATCGCATTTTAAAGGCTAGAGGG 395
Db 181 ACCTGTGCAATTAAGATATGGAGAGAGATTGTGTCAGATCGCATTTTAAAGGCTAGAGGG 240

QY 396 TCCGCAAGAAACCTTTACAGAGGAAACTATCAGGGGAGCACAAGGTTAGCGTTTTCAAA 455
Db 241 TCCGCAAGAAACCTTTACAGAGGAAACTATCAGGGGAGCACAAGGTTAGCGTTTTCAAA 300
QY 456 GAGAAAGAACTCTAG 471
Db 301 GAGAAAGAACTCTAG 316
RESULT 7
US-08-822-774-40/c
; Sequence 40, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREPE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-822-774-40
Query Match 20.5%; Score 96.4; DB 3; Length 129;
Best Local Similarity 99.0%; Pred. No. 1.2e-22;
Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCTACTTCCAGACTGGAAAATCATCAGAAAAGAAATACTTATAGAGCCATTTTCTGAGAA 60
Db 98 ATGCTACTTCCAGACTGGAAAATCATCAGAAAAGAAATACTTATAGAGCCATTTTCTGAGAA 39
QY 61 TCGCTCCACACGAGGTTATGACCTCAGAGTGGGCAG 98
Db 38 TGGCTCCACACGAGGTTATGACCTCAGAGTGGGCAG 1

RESULT 8
US-09-632-711-40/c
; Sequence 40, Application US/09632711
; Patent No. 633165
; GENERAL INFORMATION:
; APPLICANT: HOGREPE, Holly


```

; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
;
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/632,711
; FILING DATE: 04-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/822,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KOLIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
;
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (Genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
;
; US-09-632-711-40
;
; Query Match 20.5%; Score 96.4; DB 3; Length 129;
; Best Local Similarity 99.0%; Pred. No. 1.2e-22;
; Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 ATGCTACTTCAGACTCGAAATCAGAAAAGAAATCTATAGAGCCATTTCTGAAGAA 60
; Db 98 ATGCTACTTCAGACTCGAAATCAGAAAAGAAATCTATAGAGCCATTTCTGAAGAA 39
;
; QY 61 TCGCTCAACAGCAGGTTATGACCTCAGAGTGGGCAG 98
; Db 38 TGGCTCAACAGCAGGTTATGACCTCAGAGTGGGCAG 1
;
; RESULT 9
; US-09-632-703B-40/c
; Sequence 40, Application US/09632703B
; Patent No. 6379553
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF
; Protein Complexes, Isolated PEF Proteins, and Methods
; Identifying Same
;
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

;
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-09-632-702-40

Query Match 20.5%; Score 96.4; DB 3; Length 129;
Best Local Similarity 99.0%; Pred. No. 1.2e-22;
Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAA 60
Db 98 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAA 39

QY 61 TCGCTCCACCAGCAGGTTATGACCTCAGAGTGGCGAG 98
Db 38 TGGCTCCACCAGCAGGTTATGACCTCAGAGTGGCGAG 1

RESULT 11
US-09-399-003-40/c
; Sequence 40, Application US/09399003
; Patent No. 6734293
; GENERAL INFORMATION:
; APPLICANT: Hogle, Holly
; APPLICANT: Hansen, Connie J
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes
; FILE REFERENCE: 4121.0116-02
; CURRENT FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: PCT/US98/05497
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: US 08/957,709
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: US 08/822,774
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 129
; TYPE: DNA
; ORGANISM: Pyrococcus furiosus
US-09-399-003-40

Query Match 20.5%; Score 96.4; DB 4; Length 129;
Best Local Similarity 99.0%; Pred. No. 1.2e-22;
Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAA 60
Db 98 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAA 39

QY 61 TCGCTCCACCAGCAGGTTATGACCTCAGAGTGGCGAG 98
Db 38 TGGCTCCACCAGCAGGTTATGACCTCAGAGTGGCGAG 1

RESULT 12
US-08-822-774-36
; Sequence 36, Application US/08822774
; Patent No. 6183997
```

```

;
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-822-774-36

Query Match 17.4%; Score 82; DB 3; Length 84;
Best Local Similarity 97.6%; Pred. No. 7.3e-18;
Matches 82; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 87 CAGAGTGGCGAGAGGCTTTTGAAGGGGAAATTAATCGACGTGGAAAGGAGAAA 146
Db 1 CAGAGTGGCGAGAGGCTTTTGAAGGGGAAATTAATCGACGTGGAAAGGAGAAA 60

QY 147 AGTCGTTATTCTCCCAAGGAATA 170
Db 61 AGTCGTTATTCTCCCAAGGAATA 84

RESULT 13
US-09-632-711-36
; Sequence 36, Application US/09632711
; Patent No. 6333165
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,711
FILING DATE: 04-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/822,774
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-632-711-36

Query Match 17.4%; Score 82; DB 3; Length 84;
Best Local Similarity 97.6%; Pred. No. 7.3e-18;
Matches 82; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 87 CAGAGTGGCAGAGAGGCTTTTGAAGGGGAAATTAATCGACGTGAAAAGGAAGAAA 146
Db 1 CAGAGTGGCAGAGAGGCTTTTGAAGGGGAAATTAATCGACGTGAAAAGGAAGAAA 60
QY 147 AGTCGTATTCTCCCAAGGGAATA 170
Db 61 AGTCGNTATTCTCCCAAGGGAATA 84
RESULT 14
US-09-632-703B-36
Sequence 36, Application US/09632703B
Patent No. 6379553
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett, & Dunner, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: <Unknown>
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,703B
FILING DATE: 24-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/822,774
FILING DATE: 21-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: BARKER, M. Paul
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 4121.0116-04

TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 849-6613
TELEFAX: (650) 849-6666
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-632-703B-36
Query Match 17.4%; Score 82; DB 3; Length 84;
Best Local Similarity 97.6%; Pred. No. 7.3e-18;
Matches 82; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 87 CAGAGTGGCAGAGAGGCTTTTGAAGGGGAAATTAATCGACGTGAAAAGGAAGAAA 146
Db 1 CAGAGTGGCAGAGAGGCTTTTGAAGGGGAAATTAATCGACGTGAAAAGGAAGAAA 60
QY 147 AGTCGTATTCTCCCAAGGGAATA 170
Db 61 AGTCGNTATTCTCCCAAGGGAATA 84
RESULT 15
US-09-632-702-36
Sequence 36, Application US/09632702
Patent No. 6444428
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF) Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Purifying and Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/632,702
FILING DATE: 04-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/822,774
FILING DATE: 21-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-632-702-36
Query Match      17.4%; Score 82; DB 3; Length 84;
Best Local Similarity 97.6%; Pred. No. 7.3e-18;
Matches 82; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      87 CAGAGTGGGCAGAGAGGCTTTGTTAAGGGGAAATTAATCGACGTGGAAGGAAAGGAAA 146
Db      1 CAGAGTGGGCAGAGAGGCTTTGTTAAGGGGAAATTAATCGACGTGGAAGGAAAGGAAA 60

QY      147 AGTCGTTATTCTCCAAAGGGAATA 170
Db      61 AGTCGNTATTCTCCAAAGGGAATA 84
```

Search completed: July 2, 2005, 06:29:18
Job time : 147 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 14:45:07 ; Search time 420 Seconds
(without alignments)
6638.567 Million cell updates/sec

Title: US-08-957-709A-70

Perfect score: 471
Sequence: 1 atgctactccagactggaa.....caaaagagaagaactcttag 471

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	471	100.0	471	2 AAV63860	AAV63860 Polymerase
2	280.8	59.6	265118	5 AAH41227	AAH41227 Pyrococcus
C 3	256.2	54.4	110000	12 ADN46845_12	Continuation (13 o
4	256.2	54.4	110000	12 ADN47591_08	Continuation (9 of
C 5	256.2	54.4	110000	12 ADN46123_12	Continuation (13 o
6	256.2	54.4	110000	12 ADN47209_08	Continuation (9 of
C 7	256.2	54.4	110000	12 ADN46464_12	Continuation (13 o
8	256.2	54.4	110000	12 ADN47960_08	Continuation (9 of
9	49	10.4	419	6 ABX66923	ABX66923 Helicobac
C 10	44.8	9.5	3382	2 AAQ62300	AAQ62300 pBam-L3 p
11	44.2	9.4	2006	2 AAV90555	AAV90555 Nucleotid
12	44.2	9.4	2588	2 AAV90871	AAV90871 Nucleotid
13	40.2	8.5	2964	13 ADR89139	ADR89139 Nucleotid
14	39.2	8.3	597	12 ADK16664	ADK16664 Nanoarcha
15	39.2	8.3	110000	12 ADK16049_2	Continuation (3 of
16	38.6	8.2	579	12 ADJ04075	ADJ04075 DNA encod
17	38.6	8.2	89047	4 AAF28547	AAF28547 Genomic f
18	37.4	7.9	9202	1 AAN70608	AAN70608 VISA she
C 19	37.2	7.9	110021	13 ABQ32859	ABQ32859 Mouse can
20	35.4	7.5	1097	6 ABQ91119	ABQ91119 M. capsul

C 21	35	7.4	3257	2 AAT14352	AAT14352 Nk2 rece
C 22	35	7.4	3257	2 AAV38546	AAV38546 Mouse rec
23	34.8	7.4	570	4 AAI17857	AAI17857 Probe #77
24	34.8	7.4	570	4 ABA61043	ABA61043 Human foe
25	34.8	7.4	570	4 ABA62813	ABA62813 Human foe
26	34.8	7.4	570	4 AAI42828	AAI42828 Probe #11
27	34.8	7.4	570	4 AAI40941	AAI40941 Probe #96
28	34.8	7.4	570	4 ABA28956	ABA28956 Probe #74
29	34.8	7.4	570	4 ABA30099	ABA30099 Probe #85
30	34.8	7.4	570	4 AAK35225	AAK35225 Human bon
31	34.8	7.4	570	4 AAK09336	AAK09336 Human bra
32	34.8	7.4	570	4 ABS34971	ABS34971 Human liv
C 33	34.6	7.3	90435	12 ADQ59524	ADQ59524 Human can
34	34	7.2	110000	6 ABX08336_11	Continuation (12 o
35	34	7.2	110000	12 ADJ25985_11	Continuation (12 o
36	34	7.2	110000	12 ADN97989_11	Continuation (12 o
37	34	7.2	110000	12 ADQ50281_11	Continuation (12 o
C 38	33.8	7.2	349980	6 ABQ81849	ABQ81849 Bifidobac
C 39	33.6	7.1	405	2 AAX98102	AAX98102 Nucleotid
40	33.6	7.1	764	6 ABQ91833	ABQ91833 M. capsul
41	33.6	7.1	765	2 AAX98101	AAX98101 Nucleotid
42	33.6	7.1	12269	4 AAS46472	AAS46472 Tumour su
43	33.6	7.1	12269	6 ABL33411	ABL33411 Human imm
C 44	33.6	7.1	18660	4 AAK79108	AAK79108 Human imm
C 45	33.6	7.1	18671	4 AAK90763	AAK90763 Human dig

ALIGNMENTS

RESULT 1

AAV63860
ID AAV63860 standard; DNA; 471 BP.

XX AAV63860;

DT 17-OCT-2003 (revised)

DT 01-MAR-1999 (first entry)

XX Polymerase enhancing factor P45 (dUTPase) component DNA.

XX Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR; amplification; sequencing; replication; ss.

XX Pyrococcus furiosus; strain DSM 3638.

XX WO9842860-A1.

XX PD 01-OCT-1998.

XX PF 20-MAR-1998; 98WO-US005497.

XX PR 21-MAR-1997; 97US-00822774.

XX PR 24-OCT-1997; 97US-00957709.

(STRA-) STRATAGENE.

XX Hogrefe H, Hansen CJ;

XX WPI; 1998-542284/46.

XX P-PSDB; AAW72847.

XX Polymerase enhancing factor proteins, extracts and complexes - improve the polymerisation activity of nucleic acid polymerases, for use in amplification, sequencing and replication.

XX Claim 17; Page 43; 161pp; English.

XX This DNA sequence encodes the P45 component (see AAW72847) of the polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. PCR primers (see AAV63861-64) based on an N-terminal peptide (see AAW72846) of P45 were used to amplify P. furiosus genomic DNA, and sequencing was carried out on purified PCR products and plasmid mini-preps. P45 and P40

(see AAW72844) are the predominant components of PBF, which acts to enhance the activity of *P. furiosus* DNA polymerase, thereby providing replication products of greater length and purity. P45 was identified as a dUTPase, and possesses polymerase enhancing activity. The invention provides novel extracts, proteins and complexes that improve the polymerisation activity of nucleic acid polymerases, as well as DNA constructs and antibodies. Also included are methods for identifying compositions with polymerase enhancing activity, for purifying and using these compositions, and specific extracts, proteins and complexes that function to enhance polymerase activity. Nucleic acid polymerase reactions can be enhanced (claimed) by mixing a nucleic acid template, at least 1 polymerase and a composition having polymerase enhancing activity. Kits are provided for replicating nucleic acids. The kits can be used in site-directed mutagenesis, nucleic acid sequencing or amplification (preferably PCR or RT-PCR). Isolated DNA can be used in the recombinant production of P45. (Updated on 17-OCT-2003 to standardise OS field)

Sequence 471 BP; 156 A; 80 C; 124 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 471; DB 2; Length 471;
 Best Local Similarity 100.0%; Pred. No. 4.5e-142;
 Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 60
 DB 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 60

QY 61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGCGAGAGAGGCTTTTGAAGGGGAAA 120
 DB 61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGCGAGAGAGGCTTTTGAAGGGGAAA 120

QY 121 TTAATCGAGTGGAAAGGAAGAAAGTGGTTATTCCTCCAAAGGAATACGCTTAATC 180
 DB 121 TTAATCGAGTGGAAAGGAAGAAAGTGGTTATTCCTCCAAAGGAATACGCTTAATC 180

QY 181 CTAACCTCGAGAGGATAAGTTGCCCGACGATGTTATGGGGATATGAAGATAAGGAGC 240
 DB 181 CTAACCTCGAGAGGATAAGTTGCCCGACGATGTTATGGGGATATGAAGATAAGGAGC 240

QY 241 AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGGCTTGGGTTGACCCAGGATGGGATGA 300
 DB 241 AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGGCTTGGGTTGACCCAGGATGGGATGA 300

QY 301 AACTTAACTAATGCTCTACAAATGCTCCTCAATGAACCTGTGCAATTAAGATATGAGAG 360
 DB 301 AACTTAACTAATGCTCTACAAATGCTCCTCAATGAACCTGTGCAATTAAGATATGAGAG 360

QY 361 AGATTTGTCGATGCGATTTAAGGCTAGAGGGTCCGCGAAGAACCCCTTACAGAGGA 420
 DB 361 AGATTTGTCGATGCGATTTAAGGCTAGAGGGTCCGCGAAGAACCCCTTACAGAGGA 420

QY 421 AACTATCAGGGGAGCAAGGTTAGCGTTTTCAAAGAGAAAGAACTCTAG 471
 DB 421 AACTATCAGGGGAGCAAGGTTAGCGTTTTCAAAGAGAAAGAACTCTAG 471

RESULT 2

AAH41227

ID AAH41227 standard; DNA; 265118 BP.

XX AC

XX AAH41227;

XX 29-OCT-2001 (first entry)

XX DT

XX Pyrococcus abyssi genomic fragment #6.

XX DE

XX Hyperthermophilic archaeon; hyperthermophilic protein; ds.

XX KW

XX Pyrococcus abyssi.

XX OS

XX Key Location/Qualifiers

XX misc_feature 1..49980

XX FT

FT /*tag= a
 FT /note= "This sequence overlaps with the 3' end of
 XX AAH41226"

FR2792651-Al.

XX 27-OCT-2000.

XX 21-APR-1999; 99FR-00005034.

XX 21-APR-1999; 99FR-00005034.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX (IFRE-) IFREMER INST FR RECH EXPL MER.

XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

XX Querellou J, Weissenbach J, Saurin W, Heilig R;

XX PI

XX WPI; 2001-126236/14.

XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins useful in industry.

XX Claim 1; Page 593-665; 1657pp; French.

XX The present invention relates to the genomic sequence of *Pyrococcus abyssi* and *P. abyssi* proteins (see AAB96053-AAB96842). *P. abyssi* is a hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal vents. The present sequence is a fragment of the genomic sequence of *P. abyssi*. The 5' end of this sequence overlaps with the 3' end of AAH41226. The proteins of the present invention have various potential industrial uses, since the proteins are stable at very high temperatures, some up to 110 degrees centigrade. Note: This patent is in the same patent family as WO200005062, which contains additional sequences as shown in AAB99132-CC AAB99143, AAH75903-AAH75920 and AAG66436

XX Sequence 265118 BP; 75226 A; 61302 C; 54005 G; 74585 T; 0 U; 0 Other;

Query Match 59.6%; Score 280.8; DB 5; Length 265118;

Best Local Similarity 75.0%; Pred. No. 3.8e-79;

Matches 351; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 60

DB 262045 ATGCTCTTCCAGACTGGAAATCAGAAAGAAATAGAGATTTTAAAGCCATTTCTCAGAGAA 262104

QY 61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGCGAGAGAGGCTTTTGTAAAGGGGAAA 120

DB 262105 TCGCTTCAACAGCTGGGCTACGACTTAAGGGTAGGCAAGAAAGCTTACATTCAGGGGAAA 262164

QY 121 TTAATCGAGCTGGAAAGGAAGGAAAGTGGTTATTCCTCCAAAGGAATACGCTTAAATC 180

DB 262165 TTATAGATGTTGAAAGAGAGGGGCAAGTCAATATACCTTCCAAAGAAATATGCCCTAATA 262224

QY 181 CTAACCTCGAGAGGATAAGTTGCCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240

DB 262225 CTGACCTTAGAGAGGATAAGCTTCCAGATGATATTATGGGAGACATGAAGATAAGGAGC 262284

QY 241 AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGGCTTGGGTTGACCCAGGATGGGATGA 300

DB 262285 AGTCTAGCTAGGAAAGGTTCTTAGGCTCTTTTGGATGGGTAGAGCCCGGGGACGGT 262344

QY 301 AACTTAACTAATGCTCTACAAATGCTCAATGACCTGTCGAATTAAGATATGGAGAG 360

DB 262345 AATCTAACCCTTGATGCTTTTACAAATGCAATCGGAAAGAGAGGTAAATTTAAGGTACAAAGAG 262404

QY 361 AGATTTGTCGATGCGCATTTTATAAGGCTAGAGGGTCCGCGCAAGAAACCCCTTACAGAGGA 420

DB 262405 AGTTTGTGTCGATAGGCTTCTTAGGCTTGGAGCTCCCGCCCAAGAAATCCATACAGGGGC 262464

QY 421 AACTATCAGGGGAGCAAGGTTAGGTTTTCAAAGAGAAAGAACTC 468

DB 262465 AACTATCAGGAAAGCAGGAGAAATAGTCTCTCTCAAGAGATCTTAACTC 262512

RESULT 3
ADN46845_12/c

Continuation (13 of 21) of ADN46845 from base 1200001 (Thermococcus kodakaraensis KOD1 gen
WP Sequence split into 21 fragments LOCUS ADN46845 Accession Adn46845

WP	Fragment Name	Begin	End
WP	ADN46845_00	1	110000
WP	ADN46845_01	100001	210000
WP	ADN46845_02	200001	310000
WP	ADN46845_03	300001	410000
WP	ADN46845_04	400001	510000
WP	ADN46845_05	500001	610000
WP	ADN46845_06	600001	710000
WP	ADN46845_07	700001	810000
WP	ADN46845_08	800001	910000
WP	ADN46845_09	900001	1010000
WP	ADN46845_10	1000001	1110000
WP	ADN46845_11	1100001	1210000
WP	ADN46845_12	1200001	1310000
WP	ADN46845_13	1300001	1410000
WP	ADN46845_14	1400001	1510000
WP	ADN46845_15	1500001	1610000
WP	ADN46845_16	1600001	1710000
WP	ADN46845_17	1700001	1810000
WP	ADN46845_18	1800001	1910000
WP	ADN46845_19	1900001	2010000
WP	ADN46845_20	2000001	2089378

Query Match 54.4%; Score 256.2; DB 12; Length 110000;
Best Local Similarity 72.2%; Pred. NO. 2.4e-71;
Matches 333; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY	1	ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACCTATATAGAGCAATTTCTGAAGAA	60
DB	40800	ATGATGCTTCCGATTTGGAGATTAGNAGGAATTTTGTATCGAGCTTTCAGTGAAGAA	40741
QY	61	TGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGAGAGAGGCTTTTGTAAAGGGAAA	120
DB	40740	TCCTCGAGCTCGAGGATATGACCTAAGGGTTGGAGGCGAAGCTTACGCTTAATGTAAA	40681
QY	121	TTAATCGAGCTGAAAGGAAAGTTCGTTATTTCTCCAGGGAATACGCTTAATC	180
DB	40680	ATACTCGAGCTTAAGGATTACAGGGGAGTTACAAATACCCCAAGACTTACGCTTTGGTT	40621
QY	181	CTAACCTCGAGAGGATAAAGTTGCCCGAGATGTTATGGGGGATATGAAGATAAGGAGC	240
DB	40620	TTGACTCTCGAGAGGATCAAGCTTCCGGATGATGTAATGGGCGACATGAAGCTTAGGAGC	40561
QY	241	AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGGTTGACCCAGGATGGGATGGA	300
DB	40560	AGCTAGCCAGAGAGGGCTTAATCGGTTCTTTTGGCTGGGTTGACCTGGATGGATGGA	40501
QY	301	AACCTAACACTAATGCTCTCAATAGCCTCAATGAACCTGTGCAATTAAGATATGAGAG	360
DB	40500	AATCTCACTTAGCCCTGTTCACGCTTCCACGAGTCTGGAACCTCAAAATACGGCGAG	40441
QY	361	AGATTTGTGAGATCGCATTTATAAGCTAGAGGCTCCGGCAAGAAACCTTACAGAGGA	420
DB	40440	CGCTTTGTTTCAGATAGGTTTATCCGGCTGGAGGGCTGCTTAAACCCCTATCGTGA	40381
QY	421	AACCTATCAGGGAGCACAGGTTAGCGTTTTCAAAGAGAAA	461
DB	40380	AATTACCGGAGGACCAACATCTAGCGCTCTCAAAAGGAA	40340

RESULT 4
ADN47591_08

Continuation (9 of 21) of ADN47591 from base 800001 (Thermococcus kodakaraensis KOD1 gen
WP Sequence split into 21 fragments LOCUS ADN47591 Accession Adn47591

WP	Fragment Name	Begin	End
WP	ADN47591_00	1	110000
WP	ADN47591_01	100001	210000

WP	ADN47591_02	200001	310000
WP	ADN47591_03	300001	410000
WP	ADN47591_04	400001	510000
WP	ADN47591_05	500001	610000
WP	ADN47591_06	600001	710000
WP	ADN47591_07	700001	810000
WP	ADN47591_08	800001	910000
WP	ADN47591_09	900001	1010000
WP	ADN47591_10	1000001	1110000
WP	ADN47591_11	1100001	1210000
WP	ADN47591_12	1200001	1310000
WP	ADN47591_13	1300001	1410000
WP	ADN47591_14	1400001	1510000
WP	ADN47591_15	1500001	1610000
WP	ADN47591_16	1600001	1710000
WP	ADN47591_17	1700001	1810000
WP	ADN47591_18	1800001	1910000
WP	ADN47591_19	1900001	2010000
WP	ADN47591_20	2000001	2089378

Query Match 54.4%; Score 256.2; DB 12; Length 110000;
Best Local Similarity 72.2%; Pred. NO. 2.4e-71;
Matches 333; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY	1	ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACCTATATAGAGCAATTTCTGAAGAA	60
DB	48578	ATGATGCTTCCGATTTGGAGATTAGAAAGGAAATTTTGTATCGAGCTTTCAGTGAAGAA	48637
QY	61	TGCTCCAAACAGCAGGTTATGACCTCAGAGTGGGAGAGAGGCTTTTGTAAAGGGAAA	120
DB	48638	TCCTCGAGCTCGAGGATATGACCTAAGGGTTGGAGGCGAAGCTTACGCTTAATGTAAA	48697
QY	121	TTAATCGAGCTGAAAGGAAAGTTCGTTATTTCTCCAGGGAATACGCTTAATC	180
DB	48698	ATACTCGAGCTTAAGGATTACAGGGGAGTTACAAATACCCCAAGACTTACGCTTTGGTT	48757
QY	181	CTAACCTCGAGAGGATAAAGTTGCCCGAGATGTTATGGGGGATATGAAGATAAGAGC	240
DB	48758	TTGACTCTCGAGAGGATCAAGCTTCCGGATGATGTAATGGGCGACATGAAGCTTAGGAGC	48817
QY	241	AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGGTTGACCCAGGATGGGATGGA	300
DB	48818	AGCTTAGCCAGAGAGGGCTTAATCGGTTCTTTTGGCTGGGTTGACCTGGATGGATGGA	48877
QY	301	AACCTAACACTAATGCTCTCAATAGCCTCAAAATGAACCTGTGCAATTAAGATATGGAGAG	360
DB	48878	AATCTCACTTAGCCCTGTTCACGCTTCCACGAGTCTGTGGAACCTCAAAATACGGCGAG	48937
QY	361	AGATTTGTGAGATCGCATTTATAAGCTAGAGGCTCCGGCAAGAAACCTTACAGAGGA	420
DB	48938	CGCTTTGTTTCAGATAGGTTTATCCGGCTGGAGGGGCTGCTTAAACCCCTATCGTGA	48997
QY	421	AACTATCAGGGAGCACAGGTTAGCGTTTTCAAAGAGAAA	461
DB	48998	AATTACCGGAGGACCAACATCTAGCGCTCTCAAAAGGAA	49038

RESULT 5
ADN46123_12/c

Continuation (13 of 21) of ADN46123 from base 1200001 (Thermococcus kodakaraensis KOD1
WP Sequence split into 21 fragments LOCUS ADN46123 Accession Adn46123

WP	Fragment Name	Begin	End
WP	ADN46123_00	1	110000
WP	ADN46123_01	100001	210000
WP	ADN46123_02	200001	310000
WP	ADN46123_03	300001	410000
WP	ADN46123_04	400001	510000
WP	ADN46123_05	500001	610000
WP	ADN46123_06	600001	710000
WP	ADN46123_07	700001	810000
WP	ADN46123_08	800001	910000
WP	ADN46123_09	900001	1010000
WP	ADN46123_10	1000001	1110000

WP	ADN46123_11	1100001	1210000
WP	ADN46123_12	1200001	1310000
WP	ADN46123_13	1300001	1410000
WP	ADN46123_14	1400001	1510000
WP	ADN46123_15	1500001	1610000
WP	ADN46123_16	1600001	1710000
WP	ADN46123_17	1700001	1810000
WP	ADN46123_18	1800001	1910000
WP	ADN46123_19	1900001	2010000
WP	ADN46123_20	2000001	2089378

Query Match
Best Local Similarity
Matches 333; Conservative

54.4%;
72.2%;
0;

Score 256.2;
Pred. No. 2.4e-71;
Mismatches 128;

DB 12;
Indels 0;

Length 110000;
Gaps 0;

Qy	1	ATGCTACTTCCAGACTGGAAATACAAAAAGAAATACTTTATAGAGCCAATTTTTCTGAAGA	60
Dd	40800	ATGATGTCCTCCCATTGGAGAGATTAGAAAGGAAATTTTGATCGAGCCTTTCAGTGAAGA	40741
Qy	61	TCGCTCCAACCAGCAGGTTATGACCTCAGAGTGGGCAGAGAGCGTTTGTGTTAAGGGGAAA	120
Dd	40740	TCCCTGCAGCTGCAGGATATGACCTTAAGGGTTGGAGGCCAAGCTTACGTAATGSTAAA	40681
Qy	121	TTAATCGACGTGAAAGGAAGGAAAGTCGTTATTCTCCAAAGGGAATACGCCCTTAATC	180
Dd	40680	ATACTCGACGTTTAAGAGATTGACGGGGAGTTACAATACCCCCAAAGACTTACGCCCTTGTT	40621
Qy	181	CTAACCTCTCAGAGGATAAAGTTGCCCGACGATGTTTATGGGGGATATGAAGATAAGAGAC	240
Dd	40620	TTGACTCTCAGAGGATCAAGCTTCCGGATGATGTAATGGCGACATGAAGCTTAGGAGC	40561
Qy	241	AGTTTAGCAAGAAAGGGTTATGTGTTCTTTTGTGGTGTGACCCAGGATGGGATGGA	300
Dd	40560	AGCTAGCCAGAGAGGGCTTAAATCGGTTCTTTTGGCTGGGTTGACCCGTGGATGGGATGGA	40501
Qy	301	AACTTAACACTAATGCTCTACATGCTCAAATGAACCTGTCGAATTAAGATATCGAGAG	360
Dd	40500	AATCTCACTCTAGCCCTGTTTCAACGCTTCCACGATCTGTGGAACTCAAATACGCCGAG	40441
Qy	361	AGATTTGTGCAGATCGCATTTAAGGCTAGAGGGTCCGGCAGAAGAACCCCTTACAGAGA	420
Dd	40440	CGCTTTGTTCAGATAGGTTTATCCGGCTCGAGGGGCCCTGCTAAAGAACCCCTATCGTGA	40381
Qy	421	AACATACAGGGGAGCACAAAGTTAGCGTTTTCAAAGAGAAA	461
Dd	40380	AAATTACAGGGAAGCAAACTATGAGCGCTCTCAAAAGGAA	40340

RESULT 6			Continuation (9 of 21) of ADN47209 from base 800001 (Thermococcus kodakaraensis KOD1 gene)	
WP	WP	WP	WP	WP
Fragment Name	Begin	End	Accession	LOCUS
ADN47209_00	1	110000	ADN47209	ADN47209
ADN47209_01	100001	210000		
ADN47209_02	200001	310000		
ADN47209_03	300001	410000		
ADN47209_04	400001	510000		
ADN47209_05	500001	610000		
ADN47209_06	600001	710000		
ADN47209_07	700001	810000		
ADN47209_08	800001	910000		
ADN47209_09	900001	1010000		
ADN47209_10	1000001	1110000		
ADN47209_11	1100001	1210000		
ADN47209_12	1200001	1310000		
ADN47209_13	1300001	1410000		
ADN47209_14	1400001	1510000		
ADN47209_15	1500001	1610000		
ADN47209_16	1600001	1710000		
ADN47209_17	1700001	1810000		
ADN47209_18	1800001	1910000		
ADN47209_19	1900001	2010000		

WP	ADN47209_20	2000001	2089378	
	Query Match	54.4%	Score 256.2; DB 12; Length 110000;	
	Best Local Similarity	72.2%	Pred. No. 2.4e-71;	
	Matches 333; Conservative	0;	Mismatches 128; Indels 0; Gaps 0;	
Qy	1	ATGCTACTCCAGACTGGAAATACANAAGAAATACTTATACAGGCATTTTCTGAAGAA	60	
Db	48578	ATGATGCTTCCGATTGGAGATATAGAAGGAATTTTGATCGAGCCTTTTCAGTGAAGAA	48637	
Qy	61	TCGCTCCAAACACGACGGTTATGACCTCAGAGTGGCGAGAGAGCGCTTTTGTTAAGGGGAAA	120	
Db	48638	TCCTCGAGCTCGAGATATGACCTTAGGGTTGGAGCGGAAGCTTACGTAAATGGTAAA	48697	
Qy	121	TTAATCGAGCTGGAAAGGAAGAAAGTCGTTATTCTCCAAAGGAATACGCTTAAATC	180	
Db	48698	ATACTCGACGTTAAGGATTCAGGGGAGTTTACAATACCCCAAGACTTACGCCCTGGTT	48757	
Qy	181	CTAACCCCTCGAGAGGATAAAGTTGCCGAGCATGTTATGGGGATATGAAGATAAGGAC	240	
Db	48758	TTGACTCTCGAGAGGATCAAGCTTCGGATGATGTAAATGGCGCACATGAAGCTTAGGAC	48817	
Qy	241	AGTTTAGCAAGAAAGGGGTATTGGTCTCTTTTGGCTTCAGCCAGGATGGGATGGA	300	
Db	48818	AGTCTAGCCAGAGGGCTTAATCGGTTCTTTTGGCTTGGGTTGACCTGGATGGATGGA	48877	
Qy	301	AACTTAACACTAATGCTCTACAATGCTCAAAATGAACCTGCGAAATTAAGATATGGAGAG	360	
Db	48878	AATCTCACTCTAGCCCTGTTCAACGCTTCACAGCTCTGTGGAACTCAAAATACGGCGAG	48937	
Qy	361	AGATTTTGTGCAGATCGCATTTTATAAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAGGA	420	
Db	48938	CGCTTTGTTTCAGATAGCGTTTATCCGCTGGAGGGGCTCTGTAACCAACCCCTATCGTGA	48997	
Qy	421	AACTATCAGGGGAGCAAGGTTAGCGTTTTCAAAGAAA	461	
Db	48998	AAATTACAGGGAAGCAAACTCTAGGGCTCTCAAAAGGAA	49038	

Query Match	Best Local Similarity	Mismatches	Conservative	Score	DB 12;	Length
ADN46464_12/c	72.2%	0	0	256.2;	DB 12;	Length 110000;
Continuation (13 of 21) of ADN46464 from base 1200001 (Thermococcus kodakaraensis KOD1)						
WP Sequence split into 21 fragments						
WP Fragment Name	Begin	End				
ADN46464_00	1	110000				
ADN46464_01	100001	210000				
ADN46464_02	200001	310000				
ADN46464_03	300001	410000				
ADN46464_04	400001	510000				
ADN46464_05	500001	610000				
ADN46464_06	600001	710000				
ADN46464_07	700001	810000				
ADN46464_08	800001	910000				
ADN46464_09	900001	1010000				
ADN46464_10	1000001	1110000				
ADN46464_11	1100001	1210000				
ADN46464_12	1200001	1310000				
ADN46464_13	1300001	1410000				
ADN46464_14	1400001	1510000				
ADN46464_15	1500001	1610000				
ADN46464_16	1600001	1710000				
ADN46464_17	1700001	1810000				
ADN46464_18	1800001	1910000				
ADN46464_19	1900001	2010000				
ADN46464_20	2000001	2089378				
Query Match	54.4%	Score	256.2;	DB 12;	Length	110000;
Best Local Similarity	72.2%	Pred.	No. 2.4e-71;			
Mismatches 333;	Conservative	0;	Mismatches 128;	Indels	0;	Gaps 0;
Qy	1	ATGCTACTCCAGCTCGAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAA	60			
Db	40800	ATGATGCTTCCCGATCGGAAGATTAGAAAGAAATTTTGATCGAGCCCTTCAGTGAAGAA	40741			


```
QY 61 TCCTTCAACACGAGGTTATGACCTCAGAGTGGGAGAGAGGCTTTTGTAAAGGGAAA 120
DB 40740 TCCCTGCAGCCTGCAGGATATGACCTAAGGGTTGGAGGCGAAGCTTACGTTAAATGTTAAA 40681
QY 121 TTAATCGAGCTGGAAGAAGGAAAGTGGTTATTTCTCCCAAGGGAATACGCTTTAATC 180
DB 40680 ATACTCGAGTTAAGGATTCAGGGGAGTTACAAATACCCCAAGACTTACGCTTGGTT 40621
QY 181 CTAACCTTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
DB 40620 TTGACTCTCGAGAGGATCAAGCTTCGGATGATGTAATGGGCGACATGAAGCTTAGGAGC 40561
QY 241 AGTTTACGAGAGGAGGTTATGTTCTTTTGGCTTGGTTGACCCAGAGATCGGATGGA 300
DB 40560 AGTCTAGCCAGAGAGGCTTAAATCGGTTCTTTTGGCTTGGTTGACCCAGATGGATGGA 40501
QY 301 AACTTAACCTAATGCTCTACAAATGCTCAAAATGAACCTGTGCAATTAAGATATGGAGAG 360
DB 40500 AATCTCACTCTAGCCCTGTTCACAGCTTCACACGAGTCTGTGGAATCTCAAAATACGCGAG 40441
QY 361 AGATTTGTGAGATCGCATTTATAGGCTAGAGGGTCCGCGAAGAAACCCCTTACAGAGGA 420
DB 40440 CGCTTTGTTACAGATAGGTTTATCCGCTGAGAGGGCTGCTTAAAAACCCCTATCTGGA 40381
QY 421 AACTATCAGGGAGCAAGGTTAGCGTTTTCAAAGAGAAA 461
DB 40380 AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAAGGAA 40340
```

RESULT 8

ADN47960_08
Continuation (9 of 21) of ADN47960 from base 800001 (Thermococcus kodakaraensis KOD1 Gen

WP Sequence split into 21 fragments LOCUS ADN47960 Accession Adn47960

Fragment Name	Begin	End
WP ADN47960_00	1	110000
WP ADN47960_01	100001	210000
WP ADN47960_02	200001	310000
WP ADN47960_03	300001	410000
WP ADN47960_04	400001	510000
WP ADN47960_05	500001	610000
WP ADN47960_06	600001	710000
WP ADN47960_07	700001	810000
WP ADN47960_08	800001	910000
WP ADN47960_09	900001	1010000
WP ADN47960_10	1000001	1110000
WP ADN47960_11	1100001	1210000
WP ADN47960_12	1200001	1310000
WP ADN47960_13	1300001	1410000
WP ADN47960_14	1400001	1510000
WP ADN47960_15	1500001	1610000
WP ADN47960_16	1600001	1710000
WP ADN47960_17	1700001	1810000
WP ADN47960_18	1800001	1910000
WP ADN47960_19	1900001	2010000
WP ADN47960_20	2000001	2089378

Query Match 54.4%; Score 256.2; DB 12; Length 110000;

Best Local Similarity 72.2%; Pred. No. 2.4e-71; Mismatches 128; Indels 0; Gaps 0;

Matches 333; Conservative 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAAGAAATACTTATAGAGCCATTTTCTGAAGAA 60

DB 48578 ATGATGCTTCCCGATTTGGAGATTAGAAAAGGAAATTTTGTATCGAGCCTTTCAGTGAAGAA 48637

QY 61 TCCTCTCAACACGAGGTTATGACCTCAGAGTGGGAGAGAGGCTTTTGTAAAGGGAAA 120

DB 48638 TCCTCTCAGCCTGCAGATATGACCTAAGGGTTGGAGGCGAAGCTTACGTTAAATGTTAAA 48697

QY 121 TTAATCGAGCTGGAAGAAGGAAAGTGGTTATTTCTCCCAAGGGAATACGCTTTAATC 180

DB 48698 ATACTCGAGTTAAGGATTCAGGGGAGTTACAAATACCCCAAGACTTACGCTTGGTT 48757

```
QY 181 CTAACCTTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGATATGAAGATAAGGAGC 240
DB 48758 TTGACTCTCGAGAGGATCAAGCTTCGGATGATGTAATGGGCGACATGAAGCTTAGGAGC 48817
QY 241 AGTTTACGAGAGGAGGCTTATGTTCTTTTGGCTTGGTTGACCCAGAGATCGGATGGA 300
DB 48818 AGTCTAGCCAGAGAGGCTTAAATCGGTTCTTTTGGCTTGGTTGACCCAGATGGATGGA 48877
QY 301 AACTTAACTAATGCTCTACAAATGCTCAAAATGAACCTGTGCAATTAAGATATGGAGAG 360
DB 48878 AATCTCACTCTAGCCCTGTTCACAGCTTCCAACGAGTCTGTGGAATCTCAAAATACGCGAG 48937
QY 361 AGATTTGTGACAGATCGCATTTATAGGCTAGAGGGTCCGCGAAGAAACCCCTTACAGAGGA 420
DB 48938 CGCTTTGTTACAGATAGGTTTATCCGCTGAGAGGGGCTGCTTAAAAAACCCCTATCTGGA 48997
QY 421 AACTATCAGGGAGCAACAGGTTAGCGTTTTCAAAGAGAAA 461
DB 48998 AATTACCAGGAAGCAACATCTAGCGCTCTCAAAAAGGNA 49038
```

RESULT 9

ABX66923

ID ABX66923 standard; DNA; 419 BP.

XX ABX66923;

XX AC

XX DT 07-MAY-2003 (first entry)

XX DE Helicobacter pylori selected interacting domain (SID) DNA #1522.

XX KW Protein-protein interaction; ulcer; selected interacting domain; SID;

XX KW Gene; db.

XX OS Helicobacter pylori.

XX PN WO200266501-A2.

XX PD 29-AUG-2002.

XX PF 28-DEC-2001; 2001WO-BP015428.

XX PR 02-JAN-2001; 2001US-0259302P.

XX PA (HYBR-) HYBRIGENICS.

XX PA (INSP) INST PASTEUR.

XX PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;

XX WPI; 2002-674910/72.

XX DR P-PSDB; ABU52179.

XX PT New complexes of protein-protein interactions in Helicobacter pylori.

XX PT useful for identifying modulating compounds for treating or preventing

XX PT ulcers in mammals.

XX PS Claim 7; Page 457; 642pp; English.

XX CC The invention describes a complex of protein-protein interactions in

XX CC Helicobacter pylori selected from 421 complexes given in the

XX CC specification. The complex of protein-protein interactions are useful for

XX CC screening for agents which modulate the interaction of proteins.

XX CC Modulating compounds which binds to a targeted bacterial protein may be

XX CC used for treating or preventing ulcers in a human or animal. This

XX CC sequence encodes a selected interacting domain (SID), identified via

XX CC protein-protein interactions

XX SQ Sequence 419 BP; 129 A; 73 C; 99 G; 118 T; 0 U; 0 Other;

Query Match 10.4%; Score 49; DB 6; Length 419;

Best Local Similarity 48.7%; Pred. No. 3.4e-05;

Matches 133; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 112 AAGGGGAATTATCGACGTGGAAGGAAGAAAGTCTTATTCCTCCAGGGAATAC 171
 |||||
 Db 96 AACGGCACTAAATGATCGAGTAAGAGAGGTATTTTATCTTGGCGCTAACCGGTC 155
 |||||
 QY 172 GCCTTAATCCTAACCTCGAGAGGATAAAGTTGCCCGACGATGTTATGGGGATATGAAG 231
 |||||
 Db 156 GCCCTAGCCCATACGATAGATGATTTTAAATGCTTAAAGACACTTTAGCGATTTGTTTA 215
 |||||
 QY 232 ATAAGAGCAGTTTACGAGAAGGGTTATGTTCTTTTGGCTTGGTTGACCCAGA 291
 |||||
 Db 216 GGCATAAAGCATTACCTAGGTGTGGGATTATGTGAATGTTACGCTTTTGGAGCCGAA 275
 |||||
 QY 292 TGGGATGGAACCTTAACACTATGCTCTACAATGCTCAATGAACCTGTGCAATTAAGA 351
 |||||
 Db 276 TTTGAAGGCTATTATACGATTGAAATTTCTACACCACTAATTTACCGCTTAAGTCTAT 335
 |||||

RESULT 10
 AAQ62300/c
 ID AAQ62300 standard; DNA; 3382 BP.

AC AAQ62300;

DT 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 27-MAY-1994 (first entry)

DE pDam-L3 plasmid fragment encoding thermostable ligase.

XX Ligase, thermostable; thermostability; Desulfohalobus ambivalens;
 KW archaeobacteria; detection; ligation; ATP; adenosine triphosphate; ss.
 XX Acidianus ambivalens.

Key Location/Qualifiers
 CDS 1..273
 /tag= a
 /note= "Open reading frame 4."
 complement(250..271)
 CDS /tag= b
 /note= "Open reading frame 3."
 878..883
 /tag= c
 908..2707
 /tag= d
 /product= "Ligase."
 2713..2722
 /tag= e
 /note= "Potential transcription termination signal."
 2804..2811
 /tag= f
 /note= "Experimental transcription termination signal."
 2961..3382
 /tag= g
 /note= "Open reading frame 2."

DE4217134-A1.

25-NOV-1993.

23-MAY-1992; 92DE-04217134.

23-MAY-1992; 92DE-04217134.

(BOEF) BOEHRINGER MANNHEIM GMBH.
 (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

Rueger R, Kessler C, Kaletta C, Jarsch M, Kletzin A;

DR WPI; 1993-378402/48.
 DR P-PSDB; AAR44614, AAR51076, AAR51077, AAR51078.
 XX Thermostable ligase from archaeobacteria - and DNA coding for it, useful
 PT for nucleic acid detection.
 XX Claim 11; Fig 1; 26pp; German.
 XX The DNA ligase is ATP dependent and is useful for detecting nucleic acids
 CC by hybridising two oligonucleotides with adjacent sequences of the target
 CC nucleic acid, ligating the two oligonucleotides (using the ligase) and
 CC detecting the ligation product. (Updated on 25-MAR-2003 to correct PN
 CC field.) (Updated on 27-AUG-2003 to correct OS field.)

XX Sequence 3382 BP; 1188 A; 512 C; 683 G; 999 T; 0 U; 0 Other;

Query Match 9.5%; Score 44.8; DB 2; Length 3382;
 Best Local Similarity 47.9%; Pred. No. 0.002;
 Matches 161; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

QY 120 ATTAATCGACGTGGAAGGAAGAAAGTCTTATTCCTCCAGGGAATACGCTTAAT 179
 |||||
 Db 598 ATTTTATGAGATAGAGAAAGGAGATGAATTCATCTTTACCTAACGAACATGTACTGTT 539
 |||||
 QY 180 CCTAACCTCTCGAGAGGATAAAGTTGCCCGACGATGTTATGGGGGATATGAAGATAAGGAG 239
 |||||
 Db 538 AGTTACTGAGGATGTTAACTGCCGACGACGTAATGGCTTTTCGTCAACTTAAGGTC 479
 |||||
 QY 240 CAGTTTAGCAAGAGAGGGGTTATGTTCTTTCTGGTGGTTGACCCAGGATGGGATGG 299
 |||||
 Db 478 TTTCTTTGCCAGGCTGGGTCTTTTGTACCCAGCTATCGTAGATCGAGGCTTTGAAGG 419
 |||||
 QY 300 AACTTAACACTATGCTCTACAATGCTCAATGAACCTGTCGAATTAAGATATGGAGA 359
 |||||
 Db 418 ACAACTCAATAGAAAGTACTAGGTTCTGCTTTT---CCCGTGAAGATAAAGAGGGGAC 362
 |||||
 QY 360 GAGATTTGTGAGATCGCATTTATAAGGCTAGAGGGTCCGGCAAGAAACCTTTACAGAGG 419
 |||||
 Db 361 TCGCTTTCTTACCTCATTTTCGCCAGAACATTACACACAGTAGAGATCCCTTATCATGG 302
 |||||
 QY 420 AACTATCAGGGGAGCACAAAGGTTAGCGTTTTCATAA 455
 |||||
 Db 301 CAATATCAAGGACAAACAAAGGCGTAACCTTTTACCAAA 266
 |||||

RESULT 11

AAV90555

ID AAV90555 standard; DNA; 2006 BP.

AC AAV90555;

XX 20-MAR-2003 (revised)

DT 18-FEB-1999 (first entry)

XX Nucleotide sequence from clone B17.

XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.

XX Helicobacter pylori.

XX WO9849314-A2.

XX 05-NOV-1998.

XX 25-APR-1998; 98WO-US008487.

XX 25-APR-1997; 97US-0045107P.

PR 14-OCT-1997; 97US-0061958P.

XX (GENE-) GENELABS TECHNOLOGIES INC.

XX Chow TP, Fry KE, Lim MY, Mcatee CP;

XX WPI; 1999-009433/01.
 XX New Helicobacter pylori antigens and related nucleic acid sequences -
 PT useful in serological diagnosis and protective vaccines, providing long-
 PT lasting immune response.
 XX Claim 27; Page 104; 402pp; English.
 XX The present sequence encodes a Helicobacter pylori antigenic protein that
 CC is characterised by immunoreactivity with H. pylori-positive antisera.
 CC The proteins are highly immunogenic and induce a long-lasting immune
 CC response that persists even after antimicrobial treatment. In antibody-
 CC detection assays, on sera, plasma, urine, saliva etc., they are highly
 CC sensitive and specific. The specification also describes 69 previously
 CC unrecognised immunogenic cluster families. H. pylori antigens are used to
 CC detect H. pylori-specific antibodies, for diagnosing infection or to
 CC confirm eradication of infection, and in vaccines to protect against H.
 CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
 CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
 XX Sequence 2006 BP; 690 A; 377 C; 374 G; 565 T; 0 U; 0 Other;
 SQ

Query Match 9.4%; Score 44.2; DB 2; Length 2006;
 Best Local Similarity 47.8%; Pred. No. 0.0025;
 Matches 130; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 112 AAGGGGAAATTAATCGACGTGGAAAGGAAAGGAAAGTCTTATCTCCCAAGGGAATAC 171
 DB 343 AAGCGACTAAATTAATCGACGTGGAAAGGAAAGGAAAGTCTTATCTGCGGTACGCGTTC 402

QY 172 GCCTTAATCTAACCTCGAGAGGATAAAGTTCCCGCAGATGTTATGGGGATATGAAG 231
 DB 403 GCCTAGCTCATACGATAGATGATTTTAAATGCCTAAAGACACCTTAGCGATTGTTTA 462

QY 232 ATAGAGAGCGTTTAGCAAGAGAGGGGTTATGGTCTTCTTGGGTTGACCCAGGA 291
 DB 463 GCGAAAGCACTTACGCCAGGTGGGATCAITGTGAATGTTTACGCTTTTGGAGCCAGAA 522

QY 292 TGGGATGGAACTTAACACTAATGCTCACTAATGCCTCAATGAACCTGTCGAATTAAGA 351
 DB 523 TTGGAAGGGTATATCAGATTGAATTTCTAACACCAACCACTTACCGGCTAAAGTCTAT 582

QY 352 TATGGAGAGAGATTGTGCAGATCGCATTTATA 384
 DB 583 GCCAATGAGGGGATCGCGAAGTGGTGTGTTTTTA 615

RESULT 12
 AAV90871
 ID AAV90871 standard; DNA; 2588 BP.
 XX
 XX AAV90871;
 AC
 XX 20-MAR-2003 (revised)
 DT 18-FEB-1999 (first entry)
 XX
 XX Nucleotide sequence of cluster 28.
 DE
 XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
 XX Helicobacter pylori.
 OS
 XX WO9849314-A2.
 PN
 XX 05-NOV-1998.
 PD
 XX 25-APR-1998; 98WO-US008487.
 PF
 XX 25-APR-1997; 97US-0045107P.
 PR 14-OCT-1997; 97US-0061958P.
 XX

(GENE-) GENELABS TECHNOLOGIES INC.
 Chow TP, Fry KE, Lim MY, Mcatee CP;
 WPI; 1999-009433/01.
 XX New Helicobacter pylori antigens and related nucleic acid sequences -
 PT useful in serological diagnosis and protective vaccines, providing long-
 PT lasting immune response.
 XX Claim 22; Page 295; 402pp; English.
 XX The present sequence encodes a Helicobacter pylori antigenic protein that
 CC is characterised by immunoreactivity with H. pylori-positive antisera.
 CC The proteins are highly immunogenic and induce a long-lasting immune
 CC response that persists even after antimicrobial treatment. In antibody-
 CC detection assays, on sera, plasma, urine, saliva etc., they are highly
 CC sensitive and specific. The specification also describes 69 previously
 CC unrecognised immunogenic cluster families. H. pylori antigens are used to
 CC detect H. pylori-specific antibodies, for diagnosing infection or to
 CC confirm eradication of infection, and in vaccines to protect against H.
 CC pylori infection and related diseases (gastritis, peptic ulcer, gastric
 CC adenocarcinoma/lymphoma). (Updated on 20-MAR-2003 to correct PF field.)
 XX Sequence 2588 BP; 832 A; 510 C; 486 G; 760 T; 0 U; 0 Other;
 SQ

Query Match 9.4%; Score 44.2; DB 2; Length 2588;
 Best Local Similarity 47.6%; Pred. No. 0.0027;
 Matches 130; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 112 AAGGGGAAATTAATCGACGTGGAAAGGAAAGGAAAGTCTTATCTCCCAAGGGAATAC 171
 DB 926 AAGCGACTAAATTAATCGACGTGGAAAGGAAAGGAAAGTCTTATCTGCGGTACGCGTTC 985

QY 172 GCCTTAATCTAACCTCGAGAGGATAAAGTTCCCGCAGATGTTATGGGGATATGAAG 231
 DB 986 GCCTAGCTCATACGATAGATGATTTTAAATGCCTAAAGACACCTTAGCGATTGTTTA 1045

QY 232 ATAAGGAGCAGATTAGCAAGAGAGGGGTTATGGTCTTCTTGGGTTGACCCAGGA 291
 DB 1046 GCGAAAGCACTTACGCCAGGTGGGATCAITGTGAATGTTTGGCCCTTTTGGAGCCAGAA 1105

QY 292 TGGGATGGAACTTAACACTAATGCTCTCAATGCCTCAATGAACCTGTCGAATTAAGA 351
 DB 1106 TTGGAAGGGTATATCAGATTGAATTTCTAACACCACTTACCGGCTAAAGTCTAT 1165

QY 352 TATGGAGAGAGATTGTGCAGATCGCATTTATA 384
 DB 1166 GCCAATGAGGGGATCGCGAAGTGGTGTGTTTTTA 1198

RESULT 13
 ADR89139
 ID ADR89139 standard; DNA; 2964 BP.
 XX
 XX ADR89139;
 AC
 XX 18-NOV-2004 (first entry)
 DT
 XX Nucleotide sequence of a lysine decarboxylase enzyme.
 DE
 XX lysine decarboxylase; enzyme; ldc gene; lysine; cadaverine; L-lysine;
 KW lysine production; gene; ds.
 XX Methylophilus methylotrophus.
 OS
 XX Key Location/Qualifiers
 FH 684..2930
 FT /*tag= a
 FT /product= "lysine decarboxylase"
 XX
 XX FR2851575-A1.
 XX

```
PD 27-AUG-2004.
XX
XX
PF 25-FEB-2004; 2004PR-00001873.
XX
XX 25-FEB-2003; 2003JP-00047185.
XX
XX (AJIN ) AJINOMOTO CO INC.
XX
XX Hirano S, Yasueda H;
XX
XX WPI; 2004-618218/60.
XX
XX P-PSDB; ADR89140.
XX
XX Preparation of L-lysine by growing Methylobionas in which the endogenous
XX lysine decarboxylase gene has been inactivated, also new lysine
XX decarboxylase and nucleic acid encoding it.
XX
XX Claim 5; SEQ ID NO 3; 57pp; French.
XX
XX The present sequence encodes a lysine decarboxylase enzyme of
XX Methylophilus methylotrophus. The lysine decarboxylase gene is designated
XX ldc gene. The lysine decarboxylase enzyme decomposes lysine to
XX cadaverine. M. methylotrophus cells in which this gene has been reduced
XX or eliminated are useful for fermentative production of L-lysine in
XX methanol-based media. Inactivation of the enzyme increases the efficiency
XX of lysine production in Methylophilus cells.
XX
XX Sequence 2964 BP; 756 A; 784 C; 725 G; 699 T; 0 U; 0 Other;
XX
XX
XX Query Match 8.5%; Score 40.2; DB 13; Length 2964;
XX Best Local Similarity 46.3%; Pred. No. 0.058;
XX Matches 132; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
XX
XX QY 87 CAGAGTGGCCAGAGAGGCTTTTGTAAAGGGGAAATTAATCGACGTGGAAGGAGGAAA 146
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 195 CATAGTTGACCCCAAGCAATTTGACCCGAGTGGTTTGTGAGGTTCTCCGGCAAGGCTA 254
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 147 AGTCGTTATCTCCAAAGGGAATACGCTTAATCTTAACCTCGAGAGGATAAAGTTGCC 206
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 255 TTGCGTGATTCCTCCCTAACTCATTTGCACTGGCGGCACGGTAGAGTATTCGGTATCCC 314
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 207 CGACGATGTATGGGGGATATGAAGATAAGGAGCAGTATTAGCAAGAGAGGGGTTATTGG 266
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 315 TCGCTCTGTACTGACTGTATGCTCGGCAAGTCGACTTATGCGCGTTGCGGCATTATCGT 374
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 267 TTCCTTTTGGCTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAATGC 326
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 375 CAACGTCACCCCTTTTGAACCAAGAGTGGGAAGGCTATGTACACATAGAGTTTCAGCAACAC 434
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 327 CTCAAATGAACCTGTCTGAATTAAGATATGAGAGAGATTGTGCA 371
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 435 CACACCGCTACCGCCCAAAATTTATGCTGGCGAAGCTGTGCGCA 479
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX RESULT 14
XX ADK16664
XX ID ADK16664 standard; DNA; 597 BP.
XX
XX AC ADK16664;
XX
XX 06-MAY-2004 (first entry)
XX
XX Nanoarchaeum equitans cancer-associated (CA) gene #308.
XX
XX cancer-associated gene; CA gene; cancer; carcinoma; lymphoma; leukaemia;
XX ds; gene.
XX
XX Nanoarchaeum equitans.
XX
XX WO2003093434-A2.
XX
XX 13-NOV-2003.
XX
```

```
PF 01-MAY-2003; 2003WO-US013699.
XX
XX 01-MAY-2002; 2002US-0377447P.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Stetter KO, Waters E, Kretz K, Podar M, Richardson T;
XX Noordewier M;
XX
XX WPI; 2004-053041/05.
XX
XX P-PSDB; ADK16665.
XX
XX New recombinant cancer-associated genes, such as KCNJ9, useful for
XX diagnosing or treating carcinoma, e.g. breast, colon, rectal, pancreatic,
XX cervical, or skin cancers, lymphomas, or leukemia.
XX
XX Claim 5; SEQ ID NO 616; 251pp; English.
XX
XX The invention comprises then amino acid and coding sequences of cancer-
XX associated (CA) genes isolated from Nanoarchaeum equitans. The invention
XX also comprises the Nanoarchaeum equitans genome. The DNA and protein
XX sequences of the invention are useful for diagnosing and treating cancer
XX (e.g. carcinoma, lymphoma, or leukaemia). The present DNA sequence
XX represents a Nanoarchaeum equitans CA gene of the invention.
XX
XX Sequence 597 BP; 246 A; 80 C; 104 G; 167 T; 0 U; 0 Other;
XX
XX
XX Query Match 8.3%; Score 39.2; DB 12; Length 597;
XX Best Local Similarity 64.1%; Pred. No. 0.06;
XX Matches 59; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
XX
XX QY 6 ACTTCCAGACTGGAAAATCAGAAAAGAAATACTTATAGAGCCATTTTCTGAAGAATCGCT 65
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 21 AATTAAAGAGCTAATAGACAAAAGAGATAATTATTGAGCCTTTTTCCTATGAACAAAT 80
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 66 CCAACACAGAGGTTATGACCTCAGAGTGGGCA 97
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 81 CCAACACAGCATCATAGATCTAAGGCTAGGCA 112
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX RESULT 15
XX ADK16049_2
XX Continuation (3 of 5) of ADK16049 from base 200001 (Nanoarchaeum equitans genome.)
XX WP Sequence split into 5 fragments LOCUS ADK16049 Accession Adk16049
XX WP Fragment Name Begin End
XX WP ADK16049_0 1 110000
XX WP ADK16049_1 100001 210000
XX WP ADK16049_2 200001 310000
XX WP ADK16049_3 300001 410000
XX WP ADK16049_4 400001 490885
XX
XX Query Match 8.3%; Score 39.2; DB 12; Length 110000;
XX Best Local Similarity 64.1%; Pred. No. 0.6;
XX Matches 59; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
XX
XX QY 6 ACTTCCAGACTGGAAAATCAGAAAAGAAATACTTATAGAGCCATTTTCTGAAGAATCGCT 65
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 83678 AATTAAAGAGCTAATAGACAAAAGAGATAATTATTGAGCCTTTTTCCTATGAACAAAT 83737
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX QY 66 CCAACACAGAGGTTATGACCTCAGAGTGGGCA 97
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 83738 CCAACACAGCATCATAGATCTAAGGCTAGGCA 83769
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX
XX Search completed: July 1, 2005, 22:50:41
XX Job time : 424 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 2, 2005, 06:11:50 ; Search time 2574 Seconds
(without alignments)
8866.514 Million cell updates/sec

Title: US-08-957-709A-70

Perfect score: 471

Sequence: 1 atgctacttccagactgaa.....caagagaaagaactcttag 471

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pi:*
9: gb_pr:*
10: gb_ro:*
11: gb_ste:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	471	100.0	471	1	AY066005 Pyrococcus
2	471	100.0	471	1	AY443493 Pyrococcus
3	471	100.0	471	6	BD133238 Polymaras
4	471	100.0	471	6	AR535070 Sequence
5	471	100.0	13064	1	AE010292 Pyrococcus
6	374	79.4	253505	1	AP000007 Pyrococcus
7	316	67.1	740	6	BD133228 Polymaras
8	316	67.1	740	6	AR180965 Sequence
9	316	67.1	740	6	AR207520 Sequence
10	316	67.1	740	6	AR225643 Sequence
11	316	67.1	740	6	AR535060 Sequence
12	280.8	59.6	65118	6	BD430793_17 Continuation (18 o
13	280.8	59.6	265118	1	CNSPAX06 Pyrococcus
14	280.8	59.6	265118	6	AX041922 Sequence
15	96.4	20.5	129	6	BD133227 Polymaras
16	96.4	20.5	129	6	AR180964 Sequence
17	96.4	20.5	129	6	AR207519 Sequence
18	96.4	20.5	129	6	AR225642 Sequence
19	96.4	20.5	129	6	AR535059 Sequence

20	90.8	19.3	600	14	AF022221	Archaeal
21	82.8	17.6	32308	14	SVI414696	BD133226 Polymaras
22	82	17.4	84	6	BD133226	AR180963 Sequence
23	82	17.4	84	6	AR180963	AR207518 Sequence
24	82	17.4	84	6	AR225641	AR535058 Sequence
25	82	17.4	84	6	AR535058	AR535058 Sequence
26	82	17.4	84	6	AR535058	AR535058 Sequence
27	75.8	16.1	35450	14	SVI344259	AJ344259 Sulfolobu
28	57	12.1	110000	1	AE017261_01	Continuation (2 of
29	56.6	12.0	75294	7	AJ783769	AJ783769 Sulfolobu
30	54.2	11.5	7218	6	I66494	I66494 Sequence 14
31	49.6	10.5	10698	1	AE009852	AE009852 Pyrobacul
32	49	10.4	419	6	AX790579	AX790579 Sequence
33	49	10.4	11390	1	AE000554	AE000554 Helicobac
34	47.4	10.1	239340	1	AP000993	AP000993 Thermopla
35	46.4	9.9	543	6	BD180240	BD180240 Highly th
36	46.4	9.9	247910	1	AE017307	AE017307 Thermus t
37	46.4	9.9	263050	1	AP000981	AP000981 Sulfolobu
38	46	9.8	338100	1	TACID2	AL445064 Thermopla
39	45.6	9.7	14286	1	AE000747	AE000747 Aquifex a
40	45.4	9.6	11922	1	AE006657	AE006657 Sulfolobu
41	45.4	9.6	349174	1	AB063522	AB063522 Wiggleswo
42	44.8	9.5	3382	1	DALIG	X63438 D.ambivalen
43	44.8	9.5	3382	1	DALIG	X63438 D.ambivalen
44	44.2	9.4	2006	6	BD061680	BD061680 Antigenic
45	44.2	9.4	2588	6	BD061996	BD061996 Antigenic

ALIGNMENTS

RESULT 1	AY066005	471 bp	DNA	linear	BCT 28-JAN-2002
LOCUS	Pyrococcus furiosus dutpase gene, complete cds.				
DEFINITION	Pyrococcus furiosus dutpase gene, complete cds.				
ACCESSION	AY066005				
VERSION	AY066005.1	GI:18389121			
KEYWORDS	Pyrococcus furiosus				
ORGANISM	Pyrococcus furiosus				
REFERENCE	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus.				
AUTHORS	Hogrefe,H.H., Hansen,C.J., Scott,B.R. and Nielson,K.B.				
TITLE	Archaeal dutpase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (2), 596-601 (2002)				
MEDLINE	21664347				
PUBMED	11782527				
REFERENCE	2 (bases 1 to 471)				
AUTHORS	Hansen,C.J. and Hogrefe,H.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-DEC-2001) R&D, Stratagene, 11011 N. Torrey Pines Rd., La Jolla, CA 92037, USA				
FEATURES	Location/Qualifiers				
source	1..471				
CDS	/organism="Pyrococcus furiosus"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:2261"				
	1..471				
	/codon_start=1				
	/transl_table=11				
	/product="dUTPase"				
	/protein_id="AAL47572.1"				
	/db_xref="GI:18389122"				
	/translation="MLLPDWKIRKILIEPFSEESLPAGYDLRVCREAFVKGLIDV				
	EKEGKVIPPRYALILRLPDDVMDKIRSSLRGEGVIGSFADWPGDGNL				
	TLMLYNASNEPVRLYGERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL"				
ORIGIN					
Query Match	100.0%;	Score 471;	DB 1;	Length 471;	
Best Local Similarity	100.0%;	Pred. No. 4.6e-126;			
Matches 471;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	

QY 1 ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACCTTATAGAGCCATTTTCTGAAGAA 60
Db 1 ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACCTTATAGAGCCATTTTCTGAAGAA 60
QY 61 TCGCTCCCAACACGACGAGTATGACCTCAGAGTGGGCAGAGAGGCTTTTGTTAAGGGGAAA 120
Db 61 TCGCTCCCAACACGACGAGTATGACCTCAGAGTGGGCAGAGAGGCTTTTGTTAAGGGGAAA 120
QY 121 TTAATCGAGCTGGAAAAGAAAGTTCGTTATTCCTCCAAAGGGAATACGCCCTTAATC 180
Db 121 TTAATCGAGCTGGAAAAGAAAGTTCGTTATTCCTCCAAAGGGAATACGCCCTTAATC 180
QY 181 CTAACCCCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
Db 181 CTAACCCCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
QY 241 AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGGCTTGGGTTGACCCAGGATGGATGGA 300
Db 241 AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGGCTTGGGTTGACCCAGGATGGATGGA 300
QY 301 AACTTAACTAATGCTTACAAATGCTCAAAATGAACCTGTGCAATTAAGATATGAGAG 360
Db 301 AACTTAACTAATGCTTACAAATGCTCAAAATGAACCTGTGCAATTAAGATATGAGAG 360
QY 361 AGATTGTGCAGATCGCATTTATAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAGGA 420
Db 361 AGATTGTGCAGATCGCATTTATAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAGGA 420
QY 421 AACTATCAGGGGAGCACAGGTTAGCGTTTTCAAAAGAGAAAGAACTCTAG 471
Db 421 AACTATCAGGGGAGCACAGGTTAGCGTTTTCAAAAGAGAAAGAACTCTAG 471

RESULT 2
AY443493 471 bp DNA linear BCT 17-NOV-2003
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Pyrococcus woesei
Pyrococcus woesei
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
1 (bases 1 to 471)
Dabrowaki.S. and Klaer Ahring,B.
Cloning, expression, and purification of the His6-tagged
hyper-thermostable dUTPase from Pyrococcus woesei in Escherichia
coli: application in PCR
Protein Expr. Purif. 31 (1), 72-78 (2003)
22844507
12963343
2 (bases 1 to 471)
Dabrowaki.S. and Ahring,B.K.
Direct Submission
Submitted (20-OCT-2003) Biocentrum-DTU, Technical University of
Denmark, Soltofts Plads 227, Kgs. Lyngby 2800, Denmark
Location/Qualifiers
1. .471
/organism="Pyrococcus woesei"
/mol_type="genomic DNA"
/db_xref="taxon:2262"
1. .471
/codon_start=1
/transl_table=11
/product="dUTPase"
/protein_id="AA015897.1"
/db_xref="GI:38261504"
/translation="MLLPDWKIRKEILIPSPSESLQPAYDLRVGRBAFVKGKLI
DVEKGVIPPEYALITLERIKLPDDVGMKIRSSLAKEGVIGSFAWDPGDNL
TLMYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYQSTRLAFSKRKL"

QY 1 ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACCTTATAGAGCCATTTTCTGAAGAA 60
Db 1 ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACCTTATAGAGCCATTTTCTGAAGAA 60
QY 61 TCGCTCCCAACACGACGAGTATGACCTCAGAGTGGGCAGAGAGGCTTTTGTTAAGGGGAAA 120
Db 61 TCGCTCCCAACACGACGAGTATGACCTCAGAGTGGGCAGAGAGGCTTTTGTTAAGGGGAAA 120
QY 121 TTAATCGAGCTGGAAAAGAAAGTTCGTTATTCCTCCAAAGGGAATACGCCCTTAATC 180
Db 121 TTAATCGAGCTGGAAAAGAAAGTTCGTTATTCCTCCAAAGGGAATACGCCCTTAATC 180
QY 181 CTAACCCCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
Db 181 CTAACCCCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
QY 241 AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGGCTTGGGTTGACCCAGGATGGATGGA 300
Db 241 AGTTTAGCAAGAGAGGGGTTATGGTTCTTTTGGCTTGGGTTGACCCAGGATGGATGGA 300
QY 301 AACTTAACTAATGCTTACAAATGCTCAAAATGAACCTGTGCAATTAAGATATGAGAG 360
Db 301 AACTTAACTAATGCTTACAAATGCTCAAAATGAACCTGTGCAATTAAGATATGAGAG 360
QY 361 AGATTGTGCAGATCGCATTTATAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAGGA 420
Db 361 AGATTGTGCAGATCGCATTTATAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAGGA 420
QY 421 AACTATCAGGGGAGCACAGGTTAGCGTTTTCAAAAGAGAAAGAACTCTAG 471
Db 421 AACTATCAGGGGAGCACAGGTTAGCGTTTTCAAAAGAGAAAGAACTCTAG 471

RESULT 3
BD133238 471 bp DNA linear PAT 18-SEP-2002
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BD133238 1 GI:23228183
JP 2002505572-A/34.
unidentified
unidentified
unclassified.
1 (bases 1 to 471)
Hogrefe,H. and Hansen,C.J.
Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
isolated PEF proteins, and methods for purifying and identifying
Patent: JP 2002505572-A 34 19-FEB-2002;
STRATAGENE
OS Unidentified
PN JP 2002505572-A/34
PD 19-FEB-2002
PF 20-MAR-1998 JP 1998545828
PR 21-MAR-1997 US 08/822774,24-OCT-1997 US 08/957709 PI
HOLLY HOGREFE,CONNIE J HANSEN
PC C12P19/34,C12Q1/68,C12N9/12
CC Strandedness: Single;
CC Topology: Linear;
CC Polymerase enhancing factor (PEF) extracts, PEF protein CC
complexes, isolated PEF proteins, and methods for purifying and CC
identifying them
FH Key Location/Qualifiers
FT source 1. .471 /organism='Unidentified'.
FT Location/Qualifiers

CDS	70. .573 /gene="PF1985" /note="Function Code: 16.1 Conserved Hypothetical" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAL82109.1" /db_xref="GI:18894191" /translation="WHFIYMERHYTLKASKILGVTVKTLQNWQKQKIRVIRTPGGR RSIPSEIKRILGIEHEKRVGVYRSRTQKDDLRQVQAIQYAKEKGNVEILK ISSGLNEKRNSSKMWQTRKCOSSHHTSRQTNKIWIONTNTLSSPMNRDNNQRNSK RAKRRTH" 739. .1611 /gene="PF1986" 739. .1611 /gene="PF1986" /note="Function Code: 16.1 Conserved Hypothetical" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAL82110.1" /db_xref="GI:18894192" /translation="MKSITNQKSLKKILSTEKFKYVERKQLRDKLLKWNYYAAHYV DSAINQMLGLVKSRKRRGKVRPKLKKFVYVKSTLFTLKGSLVIRITIPKYY LEVINQALPYLLPFLKELFEGKIKLGLFPDPLVLFNFKTVYFKFDRWMSVDINL TNVTLANLYNFTRELYHIHRYVEVKROKIKISAWNKKLSEKLLKYSRRERNK SKDFLHLANKLIVEARERGMGIILEDLNGINKENVLRGSKGLNRKLSKWNARELQRLI EYKAKWFGVPVCLC" 1975. .2388 /gene="PF1987" 1975. .2388 /gene="PF1987" /note="Function Code: 16.1 Conserved Hypothetical" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAL82111.1" /db_xref="GI:18894193" /translation="WAIIGHNIIRVEFEKNVPVPGQVEVSLNPKIEELRLGEINLPNG KYKGLEVELTYNIISPEVKGGMIRMI VFYLPNRNEDVDNILDNWERKKLFGELVAE VYNFANLLIPFAMMITKEMGMPSPPIPRVVVRQ" 2457. .3575 /gene="PF1988" 2457. .3575 /gene="PF1988" /note="Function Code: 4.2 Cellular Processes: Cell division" /codon_start=1 /transl_table=11 /product="cell division protein ftsz homolog" /protein_id="AAL82112.1" /db_xref="GI:18894194" /translation="MLKLVENVGRVSEENKVPVQVPPQSSIDELKKIVEQIKARI YVVGAGACNTVRNMEVGTGAKIIAVNTDAQDLLKVAHQKILIGELTRGIGAG NDPKIGEEAKESRELDRALGDMVFITCGLGGTGTGAAPVIAETAKRLKMTATVS VYTLPTREGIRAKNAEXGLKLVKYSSTVIVIPNDKLEVA PKLPQMAPKVADEI LVQAQVITELTKPGLNLDNRVAVKMGCGVAMIGIGESDSKRALEAARQALNS PLVDVIGSAGALIHISGADVKEAQOIIIEYTRNVDPKPAQVIGIQLEPELEKTI RVMVITGVTSYITPEETPLETPEPSISISPEL" 3632. .3817 /gene="PF1989" 3632. .3817 /gene="PF1989" /note="Function Code: 4.6 Cellular Processes: Protein and peptide secretion" /codon_start=1 /transl_table=11 /product="protein transport protein sec61 gamma subunit" /protein_id="AAL82113.1" /db_xref="GI:18894195" /translation="WAEIQERIRHFWKESRRAPLVTKPNWATYKRAAKITGLGILI GLIGMLIRIVGILILGG" 3821. .4279	gene
CDS	3821. .4279 /gene="PF1990" /note="Function Code: 12.3 Transcription: RNA processing" /codon_start=1 /transl_table=11 /product="transcription antitermination protein nusG" /protein_id="AAL82114.1" /db_xref="GI:18894196" /translation="WACKIFAVRVTHGOETAKLIYSKVRTYNLPVAILAPSRVKG YIFVEAPNGVVDRAIRGIRHARGVLGCEVFPFKIEHLEKPAVSGLEPGDLVEIA GPFQKQAKVVKIDESKDEVVVQFIDAIPIVPTIKGDYVRLISLQKEE" 4291. .4785 /gene="PF1991" 4291. .4785 /gene="PF1991" /note="Function Code: 13.4 Translation: Ribosomal proteins: synthesis and modification; (rpl11p)" /codon_start=1 /transl_table=11 /product="LSU ribosomal protein L11p" /protein_id="AAL82115.1" /db_xref="GI:18894197" /translation="MPKQVVELVEGGKATGPPGLGPAIGPLGLNVKQVVDKINERATK DFAGQVFPKIIIVPTKQFIEVGPPTSQIILKBLGKSGSGPKHNVGNLTMEQ VIKIAKWKDQMLALTAKAAKEVIGTALSMGTVVEGKDPREVQKEIDEGVYDELFEK AEKS" 4864. .5514 /gene="PF1992" 4864. .5514 /gene="PF1992" /note="Function Code: 13.4 Translation: Ribosomal proteins: synthesis and modification; (rpl11p)" /codon_start=1 /transl_table=11 /product="LSU ribosomal protein L1p" /protein_id="AAL82116.1" /db_xref="GI:18894198" /translation="WPFDRQIKVAKVEAKARAKPRNFTQSVEVAVNLKDIDLKRPEN RFLKLEVLPHGRGKDVIAVADGAVAEARKLGIIVISSAELEBIASSPROARKLAK KYDFIABAPLMPKIGRYLGRNPNPVPVPTLDTLPIVEKLAKTVPIQIKNN PVHAPVGTETKMSDEIAENIAVLNAILGKERGESQVKSYYVKTMTGPAVKIES" 5520. .6539 /gene="PF1993" 5520. .6539 /gene="PF1993" /note="Function Code: 13.4 Translation: Ribosomal proteins: synthesis and modification; (rpl108)" /codon_start=1 /transl_table=11 /product="LSU ribosomal protein L10E" /protein_id="AAL82117.1" /db_xref="GI:18894199" /translation="MAHVAWKKEVEBELANLIKSPVVALVDVSSMPAYPLSQMRRRL IRENNGLRVSRNTLIELAIKKVAQELKPELEKLIYIEGGAGILVTMNPFLKFKF LQONQRPAPKPGAKVPKDVIPAGTSLAPGIVGQMGAMGIPARIBRGKVTIOKDT VYLKAGEBITELANILNALGIQPLEVGLDLAVVEDGIIYTPDVLAIIDSEYINMLQ KAYHAFNLNIAVPTPTIEALIQKALFNKAVAVAGYITKETSIDIIGRAIRAM LLAQOFLPVEDVDEIKELLSAQAVSQVQVEEKKKEEVEEKEEVEEALAGL SALFG" 6586. .6909 /gene="PF1994" 6586. .6909 /gene="PF1994" /note="Function Code: 13.4 Translation: Ribosomal proteins: synthesis and modification; (rpl12a)" /codon_start=1 /transl_table=11 /product="LSU ribosomal protein L12A" /protein_id="AAL82118.1" /db_xref="GI:18894200" /translation="MEYVIAALLHSVCKEINEENLKAVALQAGVEPDEARIKALVAA LEGVNIDEVIEKAMPVAVAAAAAPAGGGGEEKKEEKEEVEEVEEVEEALAGL	gene


```
gene          7014..7556
CDS           /gene="PF1995"
              /gene="PF1995"
              /note="Function Code: 16.1 Conserved Hypothetical"
              /codon_start=1
              /transl_table=11
              /product="hypothetical protein"
              /protein_id="AAU82119.1"
              /db_xref="GI:18894201"

Query Match      100.0%; Score 471; DB 1; Length 13064;
Best Local Similarity 100.0%; Pred. No. 5.1e-126;
Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 60
DB ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 7947
QY 61 TCGCTCAACACGAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTGTAAAGGGAAA 120
DB TCGCTCAACACGAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTGTAAAGGGAAA 7887
QY 121 TTAATCGAGCTGAAAGGAAGGAAGTGGTTATTCCTCCAGGGAATACGCTTAATC 180
DB TTAATCGAGCTGAAAGGAAGGAAGTGGTTATTCCTCCAGGGAATACGCTTAATC 7827
QY 181 CTAAACCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
DB CTAAACCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 7767
QY 241 AGTTTACGAGAGGAGGAGGTTATGTTCTTTTCTTTGTTGCTTCCACGAGATGGGATGA 300
DB AGTTTACGAGAGGAGGAGGTTATGTTCTTTTCTTTGTTGCTTCCACGAGATGGGATGA 7707
QY 301 AACTTAACACTAATGCTTACAAATGCTCAAAATGAACCTGTCAAAATGAATATGAGAG 360
DB AACTTAACACTAATGCTTACAAATGCTCAAAATGAACCTGTCAAAATGAATATGAGAG 7647
QY 361 AGATTTGTGCGAGATCCATTTATAAGCTAGAGGTCGGCAAGAAACCTTACAGAGGA 420
DB AGATTTGTGCGAGATCCATTTATAAGCTAGAGGTCGGCAAGAAACCTTACAGAGGA 7647
QY 421 AACTATCAGGGGAGCACAAGGTTAGCGTTTTCAAAGAGAAAGAACTCTAG 471
DB AACTATCAGGGGAGCACAAGGTTAGCGTTTTCAAAGAGAAAGAACTCTAG 7536

RESULT 6
AP000007 Pyrococcus horikoshii OT3 DNA linear BCT 27-MAY-2004
LOCUS     253505 bp
DEFINITION Pyrococcus horikoshii OT3 DNA, complete genome, 1485001-1738505 nt.
           position, section 7/7.
ACCESSION AP000007 AB009464 AB009465 AB009521 AB009522 AB009523 AB009524
           AB009525 AB009526 AB009527 AB009528 AB009529 AB009530 AB009531
           BA000001
VERSION   AP000007.1 GI:3236134
KEYWORDS  Pyrococcus horikoshii OT3
SOURCE    Pyrococcus horikoshii OT3
ORGANISM  Archaea; Euryarchaeota; Thermococci; Thermococcales;
           Thermococcaceae; Pyrococcus.
REFERENCE 1
AUTHORS   Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y.,
           Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y.,
           Sakai, M., Ogura, K., Otake, R., Nakazawa, H., Takamiya, M., Onofuku, Y.,
           Funahashi, T., Tanaka, T., Kudo, Y., Yamazaki, J., Kushiida, N.,
           Oguchi, A., Aoki, K., Nakamura, Y., Robb, T.F., Horikoshi, K.,
           Masuchi, Y., Shizuya, H. and Kikuchi, H.
           Complete sequence and gene organization of the genome of a
           hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3
           DNA Res. 5 (2), 55-76 (1998)

TITLE
JOURNAL
```

```
98344137
9679194
2 (bases 1 to 253505)
Director-General of Biotechnology Center, Tanaka, T.,
Kawarabayasi, Y. and Kikuchi, H.
Direct Submission
Submitted (11-JUN-1998) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center; 2-49-10 Nishihara, Shibuyaku, Tokyo 151-0066, Japan
(E-mail: bio@nite.go.jp, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)
On or before Mar 17, 1999 this sequence version replaced
gi:3131896, gi:3131920, gi:3131974, gi:3131982, gi:3132032,
gi:3132063, gi:3132118, gi:3132139, gi:3132184, gi:3132233,
gi:3133142.
Kawarabayasi, Y. is officially affiliated with the National
Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
305-0046, Japan.
Robb, T. F. is at the Center of Marine Biotechnology, University of
Maryland, Baltimore, MD, USA.
Horikoshi, K. is at the Japan Marine Science and Technology Center,
Yokosuka, Kanagawa 237-0061, Japan.
Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041,
Japan.
Shizuya, H. is at the California Institute of Technology, Pasadena,
CA, USA.
The other authors are at the National Institute of Technology and
Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
All the sequence with length 100 codons or more between ATG or GTG
and stop codon are defined as CDS
Homology analysis is performed by Smith-Waterman algorithm against
GenBank and GenPept release 103; EMBL release 52.0; SwissProt
release 34.0; PIR-Protein release 54.0; and OWL release 29.5.
E-mail address for comments and questions: genome@nite.go.jp
Restriction map, ORF organization, sequence alignment and more
information are available at W.W.W. site of Biotechnology Center,
URL: http://www.bio.nite.go.jp/.

FEATURES
            Location/Qualifiers
            1..253505
               /organism="Pyrococcus horikoshii OT3"
               /mol_type="genomic DNA"
               /strain="OT3"
               /db_xref="taxon:70601"
               /map="1485001-1738505 nt. position"

ORIGIN
Query Match      79.4%; Score 374; DB 1; Length 253505;
Best Local Similarity 87.2%; Pred. No. 9.9e-98;
Matches 410; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 60
DB 250855 ATGCTACTTCCGACTGGAAATCAGAAAGAAATACTTATAGAGCCATTTCTGAAGAA 250914
QY 61 TCGCTCAACACGAGGTTATGACCTCAGAGTGGGCGAGAGGCTTTTGTAAAGGGAAA 120
DB 250915 TCATTTCAACACGAGGTTATGATCTTAGAGTGGGTAGAGAGCTTTTGTAGTGAAG 250974
QY 121 TTAATCGAGCTGAAAGGAAGGAAGTGGTTATTCCTCCAGGGAATACGCTTAATC 180
DB 250975 TTAATGTAGTGGAAAGGAAGGAAGTATGTTTCTCCAGGGAATACGCTTAAT 251034
QY 181 CTAAACCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
DB 251035 CTAAACCTCGAGAGGATAAAGTTACCTGATGATGATGATGAAGATAAGGAGC 251094
QY 241 AGTTTACGAGAGGAGGAGGTTATGTTCTTTTCTTTGCTTCCACGAGATGGGATGA 300
DB 251095 AGCTTAGCTAGAGAGGAGGATTCAGGTTCTTTTTCCTGGGTGACCCAGATGGGATGG 251154
QY 301 AACTTAACACTAATGCTTACAAATGCTCAAAATGAACCTGTTCGAAATGAAGATGAGAG 360
DB 251155 AACTTAACCTAATGCTTACAAATGCTTCAAAATGAACCAAGTAGAATTAAGATACGAGAA 251214
```

```
QY 361 AGATTGTGCAGATCGCATTATTAAGCTAGAGGTCCGCAAGAAACCCCTTACAGAGGA 420
|||
Db 251215 AGTTTGTGCAGATTGCTTTATAGGCTAGAGGATCCACCAAGGAATCCCTATAGTGA 251274
|||

QY 421 AACTATCAGGGGAGCAACAGGTAGCGTTTTCAAAGAGAAAGAACTCTA 470
|||
Db 251275 AACTATCAGGAAGCAACAGATTGGTATTTTCAAGAGAAAGAACTCTA 251324
|||

RESULT 7
BD133228 740 bp DNA linear PAT 18-SEP-2002
LOCUS Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
DEFINITION isolated PEF proteins, and methods for purifying and identifying
them.
ACCESSION BD133228
VERSION BD133228.1 GI:23228173
KEYWORDS JP 2002505572-A/24.
SOURCE unidentified
ORGANISM unidentified
UNCLASSIFIED.
REFERENCE 1 (bases 1 to 740)
AUTHORS Hogrefe, H. and Hansen, C.J.
TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
JOURNAL isolated PEF proteins, and methods for purifying and identifying
Patent: JP 2002505572-A 24 19-FEB-2002;
STRATAGENE
OS Unidentified
PN JP 2002505572-A/24
PD 19-FEB-2002
PF 20-MAR-1998 JP 1998545828
PR 21-MAR-1997 US 08/822774, 24-OCT-1997 US 08/957709 PI
HOLLY HOGREFE, CONNIE J HANSEN
PC Cl2P19/34, Cl2Q1/68, Cl2N9/12
CC Strandedness: Single;
CC Topology: Linear;
CC Polymerase enhancing factor (PEF) extracts, PEF protein CC
complexes.
CC isolated PEF proteins, and methods for purifying and CC
identifying them
FH Key Location/Qualifiers
FT source 1..740
FT /organism='Unidentified'.
FEATURES
source Location/Qualifiers
1..740
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 215
|||
Db 1 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 60
|||
QY 216 TATGGGGGATATCAAGATTAAGAGCAGTTTACAGAGAGGGGTTATTGGTCTTTTTCG 275
|||
Db 61 TATGGGGGATATCAAGATTAAGAGCAGTTTACAGAGAGGGGTTATTGGTCTTTTTCG 120
|||
QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAATGCCTCAAAATGA 335
|||
Db 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAATGCCTCAAAATGA 180
|||
QY 336 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 395
|||
Db 181 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 240
|||
QY 396 TCCGCAAGAAACCCCTTACAGAGGAACTATATCAGGGGAGCACAAAGGTTAGCGTTTTCAAA 455
|||
Db 241 TCCGCAAGAAACCCCTTACAGAGGAACTATATCAGGGGAGCACAAAGGTTAGCGTTTTCAAA 300
|||

Query Match 67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 215
|||
Db 1 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 60
|||
QY 216 TATGGGGGATATCAAGATTAAGAGCAGTTTACAGAGAGGGGTTATTGGTCTTTTTCG 275
|||
Db 61 TATGGGGGATATCAAGATTAAGAGCAGTTTACAGAGAGGGGTTATTGGTCTTTTTCG 120
|||
QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAATGCCTCAAAATGA 335
|||
Db 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAATGCCTCAAAATGA 180
|||
QY 336 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 395
|||
Db 181 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 240
|||
QY 396 TCCGCAAGAAACCCCTTACAGAGGAACTATATCAGGGGAGCACAAAGGTTAGCGTTTTCAAA 455
|||
Db 241 TCCGCAAGAAACCCCTTACAGAGGAACTATATCAGGGGAGCACAAAGGTTAGCGTTTTCAAA 300
|||
```

```
QY 456 GAGAAAGAAACTCTAG 471
|||
Db 301 GAGAAAGAAACTCTAG 316
|||

RESULT 8
AR180965 740 bp DNA linear PAT 20-APR-2002
LOCUS Sequence 42 from patent US 6333165.
DEFINITION AR180965
ACCESSION AR180965
VERSION AR180965.1 GI:20222998
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 740)
AUTHORS Hogrefe, H.
TITLE Methods for identifying polymerase enhancing factor (PEF)
JOURNAL Patent: US 6333165-A 42 25-DEC-2001;
FEATURES Location/Qualifiers
1..740
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Query Match 67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 215
|||
Db 1 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 60
|||
QY 216 TATGGGGGATATGAAGATTAAGAGCAGTTTACAGAGAGGGGTTATTGGTCTTTTTCG 275
|||
Db 61 TATGGGGGATATGAAGATTAAGAGCAGTTTACAGAGAGGGGTTATTGGTCTTTTTCG 120
|||
QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAATGCCTCAAAATGA 335
|||
Db 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAATGCCTCAAAATGA 180
|||
QY 336 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 395
|||
Db 181 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 240
|||
QY 396 TCCGCAAGAAACCCCTTACAGAGGAACTATATCAGGGGAGCACAAAGGTTAGCGTTTTCAAA 455
|||
Db 241 TCCGCAAGAAACCCCTTACAGAGGAACTATATCAGGGGAGCACAAAGGTTAGCGTTTTCAAA 300
|||

QY 456 GAGAAAGAAACTCTAG 471
|||
Db 301 GAGAAAGAAACTCTAG 316
|||

RESULT 9
AR207520 740 bp DNA linear PAT 20-JUN-2002
LOCUS Sequence 42 from patent US 6379553.
DEFINITION AR207520
ACCESSION AR207520
VERSION AR207520.1 GI:21507294
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 740)
AUTHORS Hogrefe, H.
TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes,
JOURNAL isolated PEF proteins, and methods for purifying and identifying
same
Patent: US 6379553-A 42 30-APR-2002;
FEATURES Location/Qualifiers
1..740
/organism="unassigned DNA"
/mol_type="unassigned DNA"
ORIGIN
Query Match 67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 215
|||
Db 1 TCCTCAAGGAATACGCCTTAATCCTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 60
|||
QY 216 TATGGGGGATATCAAGATTAAGAGCAGTTTACAGAGAGGGGTTATTGGTCTTTTTCG 275
|||
Db 61 TATGGGGGATATCAAGATTAAGAGCAGTTTACAGAGAGGGGTTATTGGTCTTTTTCG 120
|||
QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAATGCCTCAAAATGA 335
|||
Db 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAATGCCTCAAAATGA 180
|||
QY 336 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 395
|||
Db 181 ACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCATTTATAAGGCTAGAGGG 240
|||
QY 396 TCCGCAAGAAACCCCTTACAGAGGAACTATATCAGGGGAGCACAAAGGTTAGCGTTTTCAAA 455
|||
Db 241 TCCGCAAGAAACCCCTTACAGAGGAACTATATCAGGGGAGCACAAAGGTTAGCGTTTTCAAA 300
|||
```

```

/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match      67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 TCCTCCAAGGGAATACGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCCGACGATGT 215
DB 1 TCCTCCAAGGGAATACGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCCGACGATGT 60

QY 216 TATGGGGGATATGAAGATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 275
DB 61 TATGGGGGATATGAAGATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 120

QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAATGA 335
DB 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAATGA 180

QY 336 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 395
DB 181 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 240

QY 396 TCCGGCAAGAACCCCTTACAGAGAACTATCAGGGGAGCACAAGGTTACGCTTTTCAAA 455
DB 241 TCCGGCAAGAACCCCTTACAGAGAACTATCAGGGGAGCACAAGGTTACGCTTTTCAAA 300

QY 456 GAGAAAGAACTCTAG 471
DB 301 GAGAAAGAACTCTAG 316

RESULT 10
AR225643
LOCUS      740 bp      DNA      linear      PAT 20-DEC-2002
DEFINITION Sequence 42 from patent US 644428.
ACCESSION AR225643
VERSION AR225643.1 GI:27263675
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 740)
AUTHORS Hogrefe,H. and Hansen,C.J.
TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF proteins, and methods for purifying and identifying same
JOURNAL Patent: US 644428-A 42 03-SEP-2002;
FEATURES
source
1. .740
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match      67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 TCCTCCAAGGGAATACGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCCGACGATGT 215
DB 1 TCCTCCAAGGGAATACGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCCGACGATGT 60

QY 216 TATGGGGGATATGAAGATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 275
DB 61 TATGGGGGATATGAAGATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 120

QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAATGA 335
DB 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAATGA 180

QY 336 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 395
DB 181 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 240

QY 396 TCCGGCAAGAACCCCTTACAGAGAACTATCAGGGGAGCACAAGGTTACGCTTTTCAAA 455
DB 241 TCCGGCAAGAACCCCTTACAGAGAACTATCAGGGGAGCACAAGGTTACGCTTTTCAAA 300

QY 456 GAGAAAGAACTCTAG 471
DB 301 GAGAAAGAACTCTAG 316

RESULT 11
AR353060
LOCUS      740 bp      DNA      linear      PAT 08-OCT-2004
DEFINITION Sequence 42 from patent US 6734293.
ACCESSION AR353060
VERSION AR353060.1 GI:53925880
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 740)
AUTHORS Hogrefe,H. and Hansen,C.J.
TITLE Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF protein, and methods for purifying and identifying
JOURNAL Patent: US 6734293-A 42 11-MAY-2004;
FEATURES
source
1. .740
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match      67.1%; Score 316; DB 6; Length 740;
Best Local Similarity 100.0%; Pred. No. 6.5e-81;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 TCCTCCAAGGGAATACGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCCGACGATGT 215
DB 1 TCCTCCAAGGGAATACGCTTAATCTTAACCTCGAGAGGATAAAGTTGCCCGACGATGT 60

QY 216 TATGGGGGATATGAAGATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 275
DB 61 TATGGGGGATATGAAGATAGGAGCAGTTTAGCAAGAGAGGGGTTATTGGTTCTTTTGC 120

QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAATGA 335
DB 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAAATGCCCTCAATGA 180

QY 336 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 395
DB 181 ACCTGTCGAATTAAGATATCGAGAGGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 240

QY 396 TCCGGCAAGAACCCCTTACAGAGAACTATCAGGGGAGCACAAGGTTACGCTTTTCAAA 455
DB 241 TCCGGCAAGAACCCCTTACAGAGAACTATCAGGGGAGCACAAGGTTACGCTTTTCAAA 300

QY 456 GAGAAAGAACTCTAG 471
DB 301 GAGAAAGAACTCTAG 316

Sequence split into 18 fragments LOCUS BD430793 Accession BD430793
Fragment Name Begin End
BD430793_00 1 110000
BD430793_01 100001 210000
BD430793_02 200001 310000
BD430793_03 300001 410000
BD430793_04 400001 510000
BD430793_05 500001 610000
BD430793_06 600001 710000
```

BD430793_07	700001	810000
BD430793_08	800001	910000
BD430793_09	900001	1010000
BD430793_10	1000001	1110000
BD430793_11	1100001	1210000
BD430793_12	1200001	1310000
BD430793_13	1300001	1410000
BD430793_14	1400001	1510000
BD430793_15	1500001	1610000
BD430793_16	1600001	1710000
BD430793_17	1700001	1765118

Continuation (18 of 18) of BD430793 from base 1700001 (BD430793 GENOME SEQUENCE AND POLY

Query Match	59.6%;	Score 280.8;	DB 6;	Length 65118;
Best Local Similarity	75.0%;	Pred. No. 1.3e-70;		
Matches 351;	Conservative 0;	Mismatches 117;	Indels 0;	Gaps 0;

Qy	1	ATGCTACTTCCAGACTGGAAATACAGAAAGAAATACTTATAGAGCCATTTTCTGAGGAA	60
Db	62045	ATGCTCTTCCAGACTGGAAATACAGAAAGAGATTTTAAATAAGCCATTCTCAGAGGAA	62104
Qy	61	TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGCGAGAGGCTTTTGTTAAGGGGAAA	120
Db	62105	TCGCTTCAACAGCTGGCTACGACTTAAGGTAGGCAAGGAGCTTACATTCAGGGNAAG	62164
Qy	121	TTAATCGACGTGGAAAAAGGAAGGAAAAGTCGTTATTCTCCAAGGGGAATACGCCTTAATC	180
Db	62165	TTCATAGATGTTGAAAGAGGGGCAAAAGTCATATACTCTCCAAAGAAATATGCCCTAATA	62224
Qy	181	CTAACCCCTCGAGAGGATAAAGTTCCCGCAGATGTTATGGGGGATATGAAGATAAGGAGC	240
Db	62225	CTGACCCCTAGAGAGGATAAAGCTTCCAGATGATATTATGGGAGACATGAAGATAAGGAGC	62284
Qy	241	AGTTTAGCAAGAGAGGGGTTATTGGTCTTTTTCGTTGGGTTGACCCAGGATGGGATGGA	300
Db	62285	AGTCTAGTAGGGAAGGTGTTCTTAGGCTCTTTTCATGGGTAGACCCGGGATGGGACGCT	62344
Qy	301	AACTTAACACTAATGCTCTCAATATGCTCAATATGACCTGTCGAATTTAAGATATGGAGAG	360
Db	62345	AATCTAACCTTGATGCTTTACATATGTCATCGAAAGGAGGTAAATTTTAAGGTACAAAGAG	62404
Qy	361	AGATTGTGCAGATCGCATTTTATAAGGCTAGAGGGTCGGGCAAGAAACCTTTACAGAGGA	420
Db	62405	AGGTTTGTCCAGATAGGCTTCTTAAGGCTTGAGGCTCCCGCCAAAGAAATCCATACAGGGC	62464
Qy	421	AACATACAGGGGACACAGGTTAGCGTTTTTCAAAGAGAAAGAACTC	468
Db	62465	AACATCAAGGAAGCAGAGGAATAGTCCTCTCAAAGAGATCTTAACCTC	62512

RESULT 13

CNSPAX06	265118 bp	DNA	circular	BCT 29-MAY-2003
LOCUS	Pyrococcus abyssi	complete genome; segment 6/6.		
DEFINITION	AJ248288	AL096836		
ACCESSION	AJ248288.1	GI:5458960		
VERSION	complete genome.			
KEYWORDS	Pyrococcus abyssi			
SOURCE	Pyrococcus abyssi			
ORGANISM	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus.			
REFERENCE	1 (bases 1 to 265118)			
AUTHORS	Gaspin,C., Cavaille,J., Erauso,G. and Bachelierie,J.P.			
TITLE	Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the Pyrococcus genomes			
JOURNAL	J. Mol. Biol.	297 (4),	895-906 (2000)	
MEDLINE	20202371			
PUBMED	10736225			
REFERENCE	2			
AUTHORS	Lecompte,O., Ripp,R., Puzos-Barbe,V., Duprat,S., Heilig,R., Dietrich,J., Thierry,J.C. and Poch,O.			
TITLE	Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea			

JOURNAL		Genome Res. 11 (6), 981-993 (2001)
MEDLINE		21275479
PUBMED		11381026
REFERENCE		3
AUTHORS		Cohen,G., Barbe,V., Flament,D., Galperin,M., Heilig,R., Ripp,R.,
		Lecompte,O., Prieur,D., Poch,O., Quellerou,J., Thierry,J.C., Van
		der Oost,J., Weissenbach,J., Zivanovic,Y. and Forterre,P.
		An integrated analysis of the genome of the hyperthermophilic
		archaeon Pyrococcus abyssi
		Mol. Microbiol. 47 (6), 1495-1512 (2003)
JOURNAL		22511545
MEDLINE		12622808
PUBMED		4 (bases 1 to 265118)
REFERENCE		Genoscope.
AUTHORS		Direct Submission
TITLE		Submitted (06-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL		BP 191 91006 EVRY Cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
		- Web : www.genoscope.cns.fr)
		Join(AJ248283.1:1..307150,AJ248284.1:51..293250,
		AJ248285.1:51..307120,
		AJ248286.1:51..288440,AJ248287.1:51..304290,AJ248288.1:51..265118).
COMMENT		
FEATURES		Location/Qualifiers
source		1..265118
		/organism="Pyrococcus abyssi"
		/mol_type="genomic DNA"
		/strains="Orsay"
		/db_xref="taxon:29292"
		164..1615
		/gene="bgal"
		164..1615
		/gene="bgal"
		/note="PAB2376"
		/codon_start=1
		/transl_table=11
		/product="bgal-2 beta-galactosidase (EC 3.2.1.23)
		(lactase)"
		/protein_id="CAB50447.1"
		/db_xref="GI:5458961"
		/db_xref="GOA:Q9UYG5"
		/db_xref="UniProt/TREMBL:Q9UYG5"
		/translation="NEFYGVVQSAPQEMGDPYRNLDTRSDMHWVDPYNIKNEL
		VSGDPESEGINNYELSIDRLAKSLGNAYQLTIENSRIFPCPYSEVEVERDQYG
		FIKSVIKRKEHLKDELANKREVRHYENLVKLLKNFTFTVLNQHNP1WLHDP1
		EVRNIEKARAGWDERAIVFESKFAAYVAMKFDKYDFWATFDEPMVTLGYLAP
		YVGPPEGLNPKAAKSVIMHQIIAARAYDSIKFSDKPVGVILNIIIPAYLNPDSK
		HVKAENYDLFHNRLFLKLVNKGKLDIDINGEYVKVPHLKRNDWIGNNYTREVVKYV
		EPYKELPLVTFPVEGVGYSANPNSISPDNNPTSDRGWEVYVPGIYDSTAEAYSE
		MFITENGADSKDILRPRIIVDHVKEVKRLRENGINVLGYFHNALTDYENWAMGFKI
		RFGLYEVDPITKRIIPRRKSVETIKKVKKGIE"
		1834..2283
		/note="PAB1019"
		/codon_start=1
		/transl_table=11
		/product="Predicted nucleic acid-binding protein, contains
		FIN domain"
		/protein_id="CAB50448.1"
		/db_xref="GI:5458962"
		/db_xref="UniProt/TREMBL:Q9UYG4"
		/translation="NPLPDITFDSVTLKWHSEKNKRLLEITLAKFTVLSMLTVQR
		YLTARAYLKRNVLEFEVKLDIYNVPVSEIILRASKLEANLIRKGVFLDLEIITA
		VTAIETGLLITDEPKYEPTRRFGDTPLDPLDKFLUREVELMVEKEII"
		complement(2276..3079)
		/note="PAB1341"
		/codon_start=1
		/transl_table=11
		/product="Hypothetical protein"
		/protein_id="CAB50449.1"
		/db_xref="GI:5458963"
		/db_xref="UniProt/TREMBL:Q9UYG3"
		/translation="MFENLSEPEAFKKLEEVLEFLDLPDKFVKVLCIEPHPDCC
		VIGMGTTIKLTERGIEVYACMTGYMGTLDSSLTGHELATIRREBEESKLLGVK
		KIWLNYNDITELFYSREVRKDLVRIIRKEKPGVFLPDPWLPEAHPDRHNTGFLAD
		AVAFSPLEPNFSNVDEIVGLGHQVFSFIADITYTNKPNFVFDITDVMELKUKAIRTHKSQ

```
CDS
3192..4340
/notes="PAB1020"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="CAB50450.1"
/db_xref="GI:5458964"
/db_xref="GOA:Q9UYG2"
/db_xref="UniProt/TREMBL:Q9UYG2"
/translations="MKEVSVVYKEILVKLGLEDRIETLEMKGGIIEDEFDGIRYVR
FKDSACKLRGTVIDSEYVIRPGFPHIKRIINLRSGIRIRFKBGEFVVEKVDGYNVR
VVMYKMGIGITGGICPTTIERIDFPVQEPFKDNPNNLILVGENAGPESPVLVEGP
PYKEDIFQFDVQIKTGRSLPVEERLUKABEYGINHVEFGTKVDVDELYQLI
ERLKGKRGIRIKMSDKIKVIKVPYANINDIKIGARVPELPGYFTSRISRLAF
YLAEKRIKEEFERVAKEIGLSALLQPFVESIFDVEQEDIHLELFKVRKRIETAYRWV
THPEKLGKIEIVDIEIKDGRWITPKRLYPDATNEIRELIGKAFVD"
4351..5658
/notes="PAB1021"
/codon_start=1
/transl_table=11
/product="Putative hydroxypyruvate reductase"
/protein_id="CAB50451.1"
/db_xref="GI:5458965"
/db_xref="UniProt/TREMBL:Q9UYG1"
/translations="MEDIALRLVKAIRAADPYKAVINSVRVEGNKLLVKDKFEIKG
KYVLAFGAACDMARAEILVEDVGAATVKYQKELKKIRVIEAGHPIDENSIR
GAIEGLKLEKVEEDIVFLISGGSGALFELPDGIGISLEDLKLTNELLKSGAKIHE
INTVRKHISKVGGKLAKKIKGTGIVLIIISDVVDNLEALASGPTVKDPTTFODAKRI
LELXGIDRVPSVRSRSHIEKINGIRGEAEETLKEELNHNFLINASISQANVEEATK
LGFKATWITILEGEAKDQIFGSIQVEIAKTGNFPFPPVILFQGETVYIEGEGG
KQGNQELATSTRKIKGLNALVAFDPTDGTGPTDAAGGIVDGETYEKLRRKGIDIE
KVLKENSYEALKKVGSLIFTGPTGTNVNMIATITSKQRT"
complement(5633..5655)
/notes="PAB1338"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="CAB50452.1"
/db_xref="GI:5458966"
/db_xref="UniProt/TREMBL:Q9UYG0"
/translations="MPIYEARVKVGISSNWTISKVWRDALAQLESDRIVKVLNIGE
VGEDSPFSALDIDIRIDELKMPKXDFGLKFFYVFGGELLVFIETENLLIYD
EKRFABFHVFLVINGKTVMQLARIYVGGAINMESKWEGLWIRVVKYSARTOYD
RSVVITQDKRPVSI FSDLEDIEIEEVEMNGKVRWKIRHPHINOSVTSYLIYIPDK
VOLVURLLYLKYTPSAMETIIKIADDFPTLKSEFOELMEKELSELSDBMEKQILVA
LYSGINPLELHQPLGITEXEIEBIEIYDRMIDKGLLKIVIRKIVDLTNEGRKIVNKLK
YGLVSM"
complement(6670..7425)
/notes="PAB1337"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="CAB50453.1"
/db_xref="GI:5458967"
/db_xref="UniProt/TREMBL:Q9UYF9"
/translations="MDVGTILIAVLVIGISVSAWISVSIQQMKNMQALMISRSL
EVEYTKVKRKTGRGSEAKPEAVAKQEVDRDVEEDIVRKIKQIIDEKVHNVANBAKK
KKERLMLLDVARGYALGVSEDEYNALFMRVLSELDEFKRLWLAAYPPEDKRLNQ
MTHVARTKLPVIREKRGKSGSVKLSFDEALIRITSSIKSAISLUDDLIKRGNP
ALVPLEIKLSQBLEKCKKVEDLEKRLQBLEAL"
complement(7468..9687)
/notes="PAB1336"
/codon_start=1
/transl_table=11
/product="Methyl accepting chemotaxis protein"
/protein_id="CAB50454.1"
/db_xref="GI:5458968"
/db_xref="GOA:Q9UYF8"
/db_xref="UniProt/TREMBL:Q9UYF8"
/translations="MQFRKKIILMAISAVITWAIKGGVITISFNKVKYGVIKGEIQAP
AIEQGGQLSAAEVAERKMPDFDFARVADYGTWATEVVEAYAGLDNSNETAREFLL
SBPKIKIKELNKDVSYYVFGDEEGHMYMPPDELPEGYDPRVRPWYVQAKEKGAITE
```

```

PYRDAATGQWVITYAVPVTVNGQFKVIGLDVVFVSTLLNETKKIKLGTCTYIALIDON
GMTLVHPNEEFINKUNIFQVPELKDADLDELKKGDEGYVLYTFQGIKGGYKRLKIT
GMIVATVPLSDITGILSTAINMKRHIYIQGFVGFVIYVAILVLIWYIKASNTL
APLEKLKYAAQALAEGRLVQSEYLIIRYIRBERDEICALIQAFVAKVMDLVGTINAI
KXLERLAEGDLSNGLTVEYRGELRDIIEIDRSVTETFRSIGSLVEMANDLEKRNAL
AQSVDVTRAINQVNEAIQOVSTEAORQOETINEITDGMRLVAQTSEESVRAMEEFSG
NTEVVSIANEGSKQDEALRIEDIQHMSRIEETVSKVAESNRIEETITNVTISIA
BOTNLIALNAAIEARAGBEAGRFVAVQAEIRKLAEESQOANDNIKSIIDKIDTDEIKE
AVEATKEGVSIVIGESSETLRDTIGYLANIATLLOETSERMTTVKEQIVRTQOEVDKAL
RALENLAASAEETTSABEVSSAIEQOQTAIEELRRAAQELKDMVGRMRQIVGFKV"
complement(9704..10195)
/notes="PAB1335"
/codon_start=1
/transl_table=11
/product="cheD chemotaxis protein cheD"
/protein_id="CAB50455.1"
/db_xref="GI:5458969"
/db_xref="UniProt/TREMBL:Q9UYF7"
/translations="WTRKIKVIGIDYAVKGEGIIISTYGLGSCVGITLYDRVTKVGL
LHALUPEAAARYGHRGNPAKYVDTLGLQLLKVLLKAGSKPRLEAKLFGGAOMFONIKS
DELKIGERNVQTAKEELKGLIRLVAEDTGGRGRTIYLDLSTGKVRMRKVIQGVIE
KVY"
complement(10179..10817)
/notes="PAB1334"
/codon_start=1
/transl_table=11
/product="cheC-1 chemotaxis protein cheC"
/protein_id="CAB50456.1"
/db_xref="GI:5458970"
/db_xref="UniProt/TREMBL:Q9UYF6"
/translations="MKSELYKIDIFKESNIAMSHALTSLSQMIGPIEMEADPVEIV
SRABFLUKLAERGVSXGFTVMDFVTEGLSLTILQPKHSALNISAVLMGMEFGSMBE
LDMEKSAIMEVGNILISAYTDILSNLIGEPSLSPKPAESLYDIEKELGRDLRNV
Query Match 59.6%; Score 280.8; DB 1; Length 265118;
Best Local Similarity 75.0%; Pred. No. 1.4e-70; Indels 0; Gaps 0;
Matches 351; Conservative 0; Mismatches 117;
QY 1 ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTATTATAGAGCCATTTCTGAAGAA 60
DB 262045 ATGCTCTTCCAGACTGGAAAATCAGAAAAGAGATTTTATAAAGCCATCTCAGAGGAA 262104
QY 61 TCGCTCAACACGACGCTTATGACCTCAGAGTGGGAGAGAGCGCTTTTGTATAGGGGAAA 120
DB 262105 TCGCTTCAACACGCTGGCTACGACTTAAGGAGTGGCAAGAGCTTACATTCAGGGAAG 262164
QY 121 TTAATCGAGTGGAAAAGGAGGAAAGTCGTTATTCTCCCAAGGGAATACGCTTAATC 180
DB 262165 TTAATAGATGTTGAAAAGGAGGGAAGTCAATACCTCCAAAGAAATATGCCTTAATA 262224
QY 181 CTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240
DB 262225 CTGACCCCTAGAGAGGATAAAGCTTCCAGATGATATTATGGGACATGAAGATAAGGAGC 262284
QY 241 AGTTTAGCAAGAGAGAGGGGTATTGTTCTTTTGTGGTTCACCCAGAGATGGGATGGA 300
DB 262285 AGCTAGCTAGGAGAGGCTGTTCTAGGCTCTTTTGCATGGGTAGACCCGGGATGGACGGT 262344
QY 301 AACTTAACACTAATGCTCTACATGCTCAATGACCTCAATGACCTGTCGAATTAAGATATGGAG 360
DB 262345 AATCTAACCTTGATGCTTTTACATGATCGGAAAAGAGAGGTAAATTTTAAAGGTACAAGAG 262404
QY 361 AGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGGTCCGGCAAGAAACCCCTTACAGAGGA 420
DB 262405 AGGTTTGTCCAGATACGCTTCTTAAGGCTTGAGGCTCCCGCAAGAAATCCATACAGGGGC 262464
QY 421 AACTATCAGGGGAGACACAGGTTTAGCGTTTTCACAGAGAGAAAGAACTC 468
DB 262465 AACTATCAAGGAAGCAGGAGAGATAGTCCCTCTCAAGAGAGATCTTAACTC 262512
RESULT 14
AX041922
```

LOCUS	AX041922	265118 bp	DNA	linear	PAT 23-NOV-2000
DEFINITION	Sequence 817 from Patent WO0065062.				
ACCESSION	AX041922				
VERSION	AX041922.1	GI:11340687			
KEYWORDS	Pyrococcus abyssi				
SOURCE	Pyrococcus abyssi				
ORGANISM	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus.				
REFERENCE	1				
AUTHORS	Forterre,P., Thierry,J.C., Prieur,D., Dietrich,J., Lecompte,O., Querellou,J., Weissenbach,J., Saurin,W., Heilig,R., Flament,D., Raffin,J.P., Henneke,G., Gueguen,Y. and Rolland,J.L.				
TITLE	Genome sequence and polypeptides of pyrococcus _i(abissy), fragment and uses thereof				
JOURNAL	Patent: WO 0065062-A 817 02-NOV-2000; CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (FR) ; IFREMER INSTITUT FRANCAIS DE LA RECHERCHE POUR L'EXPLOITANT; ON DE LA MER (FR)				
FEATURES	Location/Qualifiers				
source	1..265118 /organism="Pyrococcus abyssi" /mol_type="unassigned DNA" /db_xref="taxon:29292" /note="Original length of seq 1: 1765118 (for 812 sequences) New sequences : seq 001, from 0.000.001 to 0.349.980, length: 349.980 seq 813, from 0.300.001 to 0.649.980, length: 349.980 seq 814, from 0.600.001 to 0.949.980, length: 349.980 seq 815, from 0.900.001 to 1.249.980, length: 349.980 seq 816, from 1.200.001 to 1.549.980, length: 349.980 seq 817, from 1.500.001 to 1.765.118, length: 265.118"				
ORIGIN	Query Match 59.6%; Score 280.8; DB 6; Length 265118; Best Local Similarity 75.0%; Pred. No. 1.4e-70; Matches 351; Conservative 0; Mismatches 117; Indels 0; Gaps 0;				
QY	1	ATGCTACTTCCAGACTGGAAATCAGAAAAGAAATCTATTAGAGCCATTTCTGAAGAA	60		
Db	262045	ATGCTCTTCCAGACTGGAAATCAGAAAAGAAATCTATTAGAGCCATTTCTGAAGAA	262104		
QY	61	TGCTCCAAACCAGCAGGTATGACCTCAGAGTGGGCAGAGAGCTTTTGTAAAGGGGAAA	120		
Db	262105	TGCTTTCACACAGCTGGCTACGACTTAAGGTAGGCAAGGAGCTTACATTCAGGGGAAA	262164		
QY	121	TTAATCGACGTGGAAAGGAGAAAGTCGTTATTCTCCAAAGGAATACGCTTAAATC	180		
Db	262165	TTCATAGATGTTGAAAAGGAGGCAAAAGTCATAATACCTCCAAAAGAAATATGCCCTAATA	262224		
QY	181	CTAACCCTCGAGAGGATAAAGTTGCCCGACGATGTTATGGGGGATATGAAGATAAGGAGC	240		
Db	262225	CTGACCCCTAGAGAGGATAAAGCTTCCAGATGATATTATGGGAGACATGAAGATAAGGAGC	262284		
QY	241	AGTTTAGCAAGAGAAGGGTTATTGGTTCTTTTTCGTTGGGTTGACCCAGGATGGGATGGA	300		
Db	262285	AGTCTAGCTAGGGAAGGTGTTCTAGGCTCTTTTGCATGGGTAGACCCGGGATGGGACGGT	262344		
QY	301	AACCTTAACACTAATGCTCTACAAATGCCTCAAAATGAACCTGTGCAATTAAGATATGGAGAG	360		
Db	262345	AACTTAACCTTGATGCTTTTACAATGCATCGAAAAGGAGTAATTTTAAAGGTACAAAGAG	262404		
QY	361	AGATTTGTGCAGATCCGATTTATAGGCTTAGAGGGTCCGGCAAGAAACCTTTACAGAGGA	420		
Db	262405	AGGTTTGTCCAGATAGCCTTCTCAGGCTTGAGGCTCCCGCCCAAGAAATCCATACAGGGGC	262464		
QY	421	AACCTACAGGGGACGACAGGTTAGCGTTTTCAAAGAGAAAGAACTC	468		
Db	262465	AACCTATCAAGGAGGAGGAGATAGTCCTCTCAAAGAGATCTTAACTC	262512		
RESULT 15	BD133227/c				
LOCUS	BD133227	129 bp	DNA	linear	PAT 18-SEP-2002

DEFINITION	Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF proteins, and methods for purifying and identifying them.				
ACCESSION	BD133227				
VERSION	BD133227.1	GI:23228172			
KEYWORDS	JP 2002505572-A/23.				
SOURCE	unidentified				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 129)				
AUTHORS	Hogrefe,H. and Hansen,C.J.				
TITLE	Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF proteins, and methods for purifying and identifying				
JOURNAL	Patent: JP 2002505572-A 23 19-FEB-2002; STRATAGENE				
COMMENT	OS Unidentified PN JP 2002505572-A/23 PD 19-FEB-2002 PF 20-MAR-1998 JP 1998545828 PR 21-MAR-1997 US 08/822774, 24-OCT-1997 US 08/957709 PI PC C12P19/34,C12Q1/68,C12N9/12 HOLLY HOGREFE,CONNIE J HANSEN CC Strandedness: Single; CC Topology: Linear; CC Polymerase enhancing factor (PEF) extracts, PEF protein CC complexes, CC isolated PEF proteins, and methods for purifying and CC identifying them FH Key Location/Qualifiers FT source 1..129 FT /organism='Unidentified'. FEATURES source Location/Qualifiers 1..129 /organism="unidentified" /mol_type="genomic DNA" /db_xref="taxon:32644"				
ORIGIN	Query Match 20.5%; Score 96.4; DB 6; Length 129; Best Local Similarity 99.0%; Pred. No. 5.6e-17; Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY	1	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTATAGAGCCATTTCTGAAGAA	60		
Db	98	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTATAGAGCCATTTCTGAAGAA	39		
QY	61	TGCTCCAAACCAGCAGGTTATGACCTCAGAGTGGGCAG	98		
Db	38	TGCTCCAAACCAGCAGGTTATGACCTCAGAGTGGGCAG	1		
Search completed: July 2, 2005, 07:12:21 Job time : 2578 secs					

DEFINITION	Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF proteins, and methods for purifying and identifying them.									
ACCESSION	BD133227									
VERSION	BD133227.1 GI:23228172									
KEYWORDS	JP 2002505572-A/23.									
SOURCE	unidentified									
ORGANISM	unidentified									
REFERENCE	1 (bases 1 to 129)									
AUTHORS	Hogrefe,H. and Hansen,C.J.									
TITLE	Polymerase enhancing factor (PEF) extracts, PEF protein complexes, isolated PEF proteins, and methods for purifying and identifying									
JOURNAL	Patent: JP 2002505572-A 23 19-FEB-2002; STRATAGENE									
COMMENT	OS Unidentified PN JP 2002505572-A/23 PD 19-FEB-2002 PF 20-MAR-1998 JP 1998545828 PR 21-MAR-1997 US 08/822774,24-OCT-1997 US 08/957709 P1 PC C12P19/34,C12Q1/68,C12N9/12 CC HOLLY HOGREFE,CONNIE J HANSEN CC Strandedness: Single; CC Topology: Linear; CC Polymerase enhancing factor (PEF) extracts, PEF protein CC complexes, CC isolated PEF proteins, and methods for purifying and CC identifying them									
FH	Key Location/Qualifiers									
FT	source 1..129									
FT	/organism='Unidentified'.									
FEATURES	Location/Qualifiers									
source	1..129 /organism='unidentified' /mol_type='genomic DNA' /db_xref='taxon:32644'									
ORIGIN	Query Match 20.5%; Score 96.4; DB 6; Length 129; Best Local Similarity 99.0%; Pred.No. 5.6e-17; Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Qy	1	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACCTTATAGAGCCATTTCTGAAGAA							60	
Db	98	ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACCTTATAGAGCCATTTCTGAAGAA							39	
Qy	61	TCGCTCCCAACCAAGCAGGTTATGACCTCAGAGTGGGCAG							98	
Db	38	TGGCTCCCAACCAAGCAGGTTATGACCTCAGAGTGGGCAG							1	
Search completed: July 2, 2005, 07:12:21 Job time : 2578 secs										

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:36:20 ; Search time 30.9767 Seconds
(without alignments)
1942.879 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 156
Sequence: 1 MLLPWKIRKELIEPFSEE.....PYRGNVQSTRLAFKXKL 156

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1722976 seqs, 385795295 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1722976

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	5.1	194	15 US-10-425-114-67700	Sequence 67700, A
2	8	5.1	459	17 US-10-732-923-20236	Sequence 20236, A
3	8	5.1	459	17 US-10-732-923-20237	Sequence 20237, A
4	8	5.1	490	15 US-10-369-493-15077	Sequence 15077, A
5	8	5.1	499	15 US-10-369-493-11671	Sequence 11671, A
6	8	5.1	499	15 US-10-369-493-14468	Sequence 14468, A
7	7	4.5	45	10 US-09-764-891-3318	Sequence 3318, Ap
8	7	4.5	66	16 US-10-437-963-187764	Sequence 187764, A
9	7	4.5	71	16 US-10-425-115-360816	Sequence 360816, A
10	7	4.5	174	16 US-10-767-701-56940	Sequence 56940, A
11	7	4.5	198	16 US-10-437-963-184042	Sequence 184042, A

12	7	4.5	215	17 US-10-732-923-21372	Sequence 21372, A
13	7	4.5	220	16 US-10-437-963-147587	Sequence 147587, A
14	7	4.5	237	10 US-09-461-580A-13	Sequence 13, Appl
15	7	4.5	241	15 US-10-425-114-52925	Sequence 52925, A
16	7	4.5	250	11 US-09-809-665A-83	Sequence 83, Appl
17	7	4.5	250	17 US-10-854-299-83	Sequence 83, Appl
18	7	4.5	288	14 US-10-128-714-3054	Sequence 3054, Ap
19	7	4.5	288	14 US-10-128-714-8054	Sequence 8054, Ap
20	7	4.5	299	14 US-10-219-810-36	Sequence 36, Appl
21	7	4.5	330	16 US-10-437-963-173910	Sequence 173910, A
22	7	4.5	345	14 US-10-279-029-95	Sequence 22, Appl
23	7	4.5	345	14 US-10-219-810-22	Sequence 2575, Ap
24	7	4.5	399	15 US-10-369-493-2375	Sequence 19715, A
25	7	4.5	429	15 US-10-369-493-13715	Sequence 20285, A
26	7	4.5	448	17 US-10-732-923-20265	Sequence 9303, Ap
27	7	4.5	466	14 US-10-156-701-9303	Sequence 44737, A
28	7	4.5	501	16 US-10-767-701-44737	Sequence 70654, A
29	7	4.5	518	15 US-10-282-122A-70654	Sequence 2, Appl
30	7	4.5	592	13 US-10-047-593-2	Sequence 4, Appl
31	7	4.5	592	13 US-10-047-593-4	Sequence 10, Appl
32	7	4.5	607	14 US-10-318-780-10	Sequence 11, Appl
33	7	4.5	609	14 US-10-318-780-11	Sequence 18, Appl
34	7	4.5	609	18 US-10-220-665B-18	Sequence 905, App
35	7	4.5	661	15 US-10-369-493-905	Sequence 11079, A
36	7	4.5	678	14 US-10-156-761-11079	Sequence 155249, A
37	7	4.5	706	16 US-10-437-963-135249	Sequence 8334, Ap
38	7	4.5	874	17 US-10-732-923-8334	Sequence 158, App
39	7	4.5	1017	16 US-10-408-765A-158	Sequence 2452, Ap
40	7	4.5	1084	17 US-10-472-928-2452	Sequence 45050, A
41	7	4.5	1106	15 US-09-970-711-28	Sequence 28, Appl
42	7	4.5	1121	9 US-10-408-765A-1309	Sequence 1309, Ap
43	7	4.5	1514	16 US-10-437-963-188596	Sequence 188596, A
44	7	4.5	1688	16 US-10-437-963-146169	Sequence 146169, A
45	7	4.5	2187	16 US-10-437-963-146169	Sequence 146169, A

ALIGNMENTS

RESULT 1
US-10-425-114-67700
; Sequence 67700, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 67700
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3592-094-F1_FU1.pep
US-10-425-114-67700

Query Match 5.1% Score 8; DB 15; Length 194;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 DLRVGREA 35
| | | | |
DB 139 DLRVGREA 146

```
RESULT 2
US-10-732-923-20236
; Sequence 20236, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 20236
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-20236

Query Match          5.1%; Score 8; DB 17; Length 459;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      61  LTLERIKL 68
Db      212 LTLERIKL 219

RESULT 3
US-10-732-923-20237
; Sequence 20237, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 20237
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-20237

Query Match          5.1%; Score 8; DB 17; Length 459;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      61  LTLERIKL 68
Db      212 LTLERIKL 219

RESULT 4
US-10-369-493-15077
; Sequence 15077, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15077
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-15077

Query Match          5.1%; Score 8; DB 15; Length 499;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      25  AGYDLRVG 32
Db      454 AGYDLRVG 461

RESULT 5
US-10-369-493-11671
; Sequence 11671, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11671
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11671

Query Match          5.1%; Score 8; DB 15; Length 499;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      25  AGYDLRVG 32
Db      454 AGYDLRVG 461

RESULT 6
US-10-369-493-14468
; Sequence 14468, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14468
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14468
```



```
Query Match          5.1%; Score 8; DB 15; Length 499;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      25 AGYDLRVG 32
      |||||
Db      454 AGYDLRVG 461

RESULT 7
US-09-764-891-3318
; Sequence 3318, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3318
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-3318

Query Match          4.5%; Score 7; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SLAREGV 87
      |||||
Db      29 SLAREGV 35

RESULT 8
US-10-437-963-187764
; Sequence 187764, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 187764
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(174)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_84435C.1.pep
US-10-437-963-187764

Query Match          4.5%; Score 7; DB 16; Length 66;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      81 SLAREGV 87
      |||||
Db      5 SLAREGV 11

RESULT 9
US-10-425-115-360816
; Sequence 360816, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 360816
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_92246C.1.pep
US-10-425-115-360816

Query Match          4.5%; Score 7; DB 16; Length 71;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      128 IRLEGPA 134
      |||||
Db      1 IRLEGPA 7

RESULT 10
US-10-767-701-56940
; Sequence 56940, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 56940
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(174)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: 30951232.pep
US-10-767-701-56940

Query Match          4.5%; Score 7; DB 16; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      57 YALILTL 63
      |||||
Db      94 YALILTL 100

RESULT 11
US-10-437-963-184042
; Sequence 184042, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

```
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 184042
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(198)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_81073C.1.pap
US-10-437-963-184042

Query Match          4.5%; Score 7; DB 16; Length 198;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RLEGPAR 135
DB 181 RLEGPAR 187

RESULT 12
US-10-732-923-21372
; Sequence 21372, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 21372
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Sulfolobus solfataricus
US-10-732-923-21372

Query Match          4.5%; Score 7; DB 17; Length 215;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 RIKLPDD 71
DB 24 RIKLPDD 30

RESULT 13
US-10-437-963-147587
; Sequence 147587, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
```

```
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147587
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(220)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_48101C.1.pap
US-10-437-963-147587

Query Match          4.5%; Score 7; DB 16; Length 220;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LRVGREA 35
DB 103 LRVGREA 109

RESULT 14
US-09-461-580A-13
; Sequence 13, Application US/09461580A
; Publication No. US20030207325A1
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard
; APPLICANT: Imai, Shin-ichiro
; APPLICANT: Armstrong, Christopher
; TITLE OF INVENTION: METHODS FOR IDENTIFYING AGENTS WHICH
; TITLE OF INVENTION: ALTER HISTONE PROTEIN ACETYLATION, DECREASE AGING, INCREASE
; TITLE OF INVENTION: LIFESPAN
; FILE REFERENCE: 0050.1618-000
; CURRENT APPLICATION NUMBER: US/09/461,580A
; CURRENT FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-461-580A-13

Query Match          4.5%; Score 7; DB 10; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 EGPARNP 137
DB 43 EGPARNP 49

RESULT 15
US-10-425-114-52925
; Sequence 52925, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
```

FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 52925
LENGTH: 241
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB3060-025-H5_FLI.pep
US-10-425-114-52925

Query Match 4.5%; Score 7; DB 15; Length 241;
Best Local Similarity 100.0%; Fred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LRVGREA 35
|||
Db 187 LRVGREA 193

Search completed: July 2, 2005, 12:40:44
Job time : 30.9767 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:31:19 ; Search time 33.7674 Seconds
(without alignments)
1786.769 Million cell updates/sec

Title: US-08-957-709A-71
Perfect score: 156
Sequence: 1 MLPLDWKIRKEILIEPFSEE.....PYRGNVQSTRLAFSKKKL 156

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003Bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	156	100.0	156	2 AAW72847	Aaw72847 Polymerase
2	22	14.1	154	4 AAB96424	Aab96424 Putative
3	14	9.0	15	2 AAW72845	Aaw72845 Polymerase
4	11	7.1	14	2 AAW72849	Aaw72849 Pyrococcus
5	8	5.1	8	2 AAW72846	Aaw72846 Polymerase
6	8	5.1	490	8 ADS26044	Ads26044 Bacterial
7	8	5.1	499	8 ADS2638	Ads2638 Bacterial
8	8	5.1	499	8 ADS25435	Ads25435 Bacterial
9	7	4.5	15	2 AAW72870	Aaw72870 Polymerase
10	7	4.5	45	4 AAUI8829	AAui8829 Novel pro
11	7	4.5	45	4 AAM94660	Aam94660 Human rep
12	7	4.5	74	5 ABP07719	Abp07719 Human ORF
13	7	4.5	83	4 AAUS8795	Aaus8795 Propionib
14	7	4.5	83	6 ABMS5314	Abms5314 Propionib
15	7	4.5	102	6 ABP75762	Abp75762 Human sec
16	7	4.5	109	4 AAO06916	Aao06916 Human pol
17	7	4.5	153	3 AAB14140	Aab14140 Bordetell
18	7	4.5	162	2 AAR96207	Aar96207 Attachmen
19	7	4.5	164	3 AAG46155	Aag46155 Arabidops
20	7	4.5	164	4 AAG82692	Aag82692 S. epider
21	7	4.5	167	5 ABP39663	Abp39663 Staphyloc
22	7	4.5	167	8 ADS05751	Ads05751 Staphyloc
23	7	4.5	171	3 AAG46154	Aag46154 Arabidops
24	7	4.5	178	6 ABP98199	Abp98199 Amino aci
25	7	4.5	180	4 ABB58191	Abb58191 Drosophil

26	7	4.5	182	6 ABP98200	Abp98200 Amino aci
27	7	4.5	186	7 ABO68216	Abc68216 Pseudomon
28	7	4.5	198	3 AAG46153	Aag46153 Arabidops
29	7	4.5	240	6 ADA48536	Ada48536 Rice prot
30	7	4.5	250	3 AAB44561	Aab44561 Virulence
31	7	4.5	250	5 AAB54513	Abp54513 Pasteurel
32	7	4.5	288	6 ABJ25996	Abj25996 Aspergill
33	7	4.5	288	6 ABJ25396	Abj25396 Aspergill
34	7	4.5	299	6 ABP96392	Abp96392 Gossypium
35	7	4.5	345	6 ABP96378	Abp96378 G. hirsut
36	7	4.5	345	7 ADD19145	Add19145 Gossypium
37	7	4.5	345	7 ABO64328	Abc64328 Klebsiell
38	7	4.5	362	3 AAY70024	Aay70024 Human Pro
39	7	4.5	367	7 ABO60852	Abc60852 Klebsiell
40	7	4.5	399	8 ADNI19922	Adni19922 Bacterial
41	7	4.5	413	4 AAU29772	Aau29772 Novel hum
42	7	4.5	419	8 ABM80267	Abm80267 Tumour-as
43	7	4.5	429	8 ADS30682	Ads30682 Bacterial
44	7	4.5	443	5 ABP38922	Abp38922 Staphyloc
45	7	4.5	443	8 ADS06651	Ads06651 Staphyloc

ALIGNMENTS

RESULT 1
AAW72847
ID AAW72847 standard; protein; 156 AA.
XX
AC AAW72847;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P45 (dUTPase) component.
XX
KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication.
XX
OS Pyrococcus furiosus; strain DSM 3638.
XX
FH Key Location/Qualifiers
FT Peptide 2..15
FT /note= "N-terminal peptide used to generate primers"
XX
PN W09842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US005497.
XX
PR 21-MAR-1997; 97US-00822774.
PR 24-OCT-1997; 97US-00957709.
XX
PA (STRA-) STRATAGENE.
XX
PI Hogrefe H, Hansen CJ;
DR WPI: 1998-542284/46.
DR N-PSDB; AAV63860.
XX
PT Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX
PS Claim 17; Page 43; 161pp; English.
XX
CC This is the amino acid sequence of the P45 component of the polymerase
CC enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. The sequence is
CC predicted from a DNA sequence (see AAV63860) obtained from genomic DNA by
CC PCR. P45 and P50 (see AAW72844) are the predominant components of PEF,
CC which acts to enhance the activity of P. furiosus DNA polymerase. P45
CC functions as a dUTPase, and can be used to enhance nucleic acid

CC replication, polymerisation or PCR reactions. The invention provides
CC novel extracts, proteins and complexes that improve the polymerisation
CC activity of nucleic acid polymerases, as well as DNA constructs and
CC antibodies. Also included are methods for identifying compositions with
CC PEF activity, for purifying and using these compositions, and specific
CC extracts, proteins and complexes that function to enhance polymerase
CC activity. Nucleic acid polymerase reactions can be enhanced (claimed) by
CC mixing a nucleic acid template, at least 1 polymerase and a composition
CC having polymerase enhancing activity. Kits are provided for replicating
CC nucleic acids, for site-directed mutagenesis, for nucleic acid sequencing
CC or amplification (preferably PCR or RT-PCR). (Updated on 17-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 156 AA;

Query Match 100.0%; Score 156; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.1e-149;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLLPDWKIRKEILIEPSESLOPAGYDLRVGREAFVKGKLDVKEGKVIPPPEYALI 60
DB 1 MLLPDWKIRKEILIEPSESLOPAGYDLRVGREAFVKGKLDVKEGKVIPPPEYALI 60
QY 61 LTLERIKLPDDVGMGMKIRSSLAREGVIGSFVWDPGWDGNLTMLYNASNEPVELRYGE 120
DB 61 LTLERIKLPDDVGMGMKIRSSLAREGVIGSFVWDPGWDGNLTMLYNASNEPVELRYGE 120
QY 121 RFVQIAFIRLEGPARNPYRGNGVQSTRLAFSKRKL 156
DB 121 RFVQIAFIRLEGPARNPYRGNGVQSTRLAFSKRKL 156

RESULT 2
AAB96424
ID AAB96424 standard; protein; 154 AA.
XX
AC AAB96424;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative P. abyssi deoxycytidine deaminase.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
PN FR2792651-A1.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99FR-00005034.
XX
PR 21-APR-1999; 99FR-00005034.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
WPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode proteins
PT useful in industry.
PT
XX
PS Claim 7; Page 1118; 1657pp; French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
CC abyssi (see AAF86431 and AA41223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to

CC 110 degrees centigrade. Note: This patent is in the same patent family as
CC WO200005062, which contains additional sequences as shown in AAB99132-
CC AAB99143, AAH75903-AAH75920 and AAG66436
XX
SQ Sequence 154 AA;
Query Match 14.1%; Score 22; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 9.4e-14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 89 GSFAWVDFGWDGNLTMLYNAS 110
DB 89 GSFAWVDFGWDGNLTMLYNAS 110
RESULT 3
AAW72845
ID AAW72845 standard; peptide; 15 AA.
XX
AC AAW72845;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P45 component N-terminal peptide.
XX
KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication.
XX
OS Pyrococcus furiosus; strain DSM 3638.
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /label= Gly, Ala, Ile
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US005497.
XX
PR 21-MAR-1997; 97US-00822774.
PR 24-OCT-1997; 97US-00957709.
XX
PA (STRA-) STRATAGENE.
XX
PI Hogrefe H, Hansen CJ;
XX
WPI; 1998-542284/46.
XX
XX Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX
PS Claim 17; Page 33; 161pp; English.
XX
XX This peptide comprises an N-terminal peptide of the P45 component of the
CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. It has
CC been obtained from isolated P45; a full-length sequence is provided in
CC AAW72847. PEF, the predominant components of which are P45 and P50 (see
CC AAW72844) proteins, enhances the activity of P. furiosus DNA polymerase,
CC thereby providing replication products of greater length and purity. P45
CC is a dUTPase and possesses polymerase enhancing activity. The invention
CC provides novel extracts, proteins and complexes that improve the
CC polymerisation activity of nucleic acid polymerases. These complexes may
CC include proteins including the P45 N-terminal peptide. Also included are
CC methods for identifying compositions with polymerase enhancing activity,
CC for purifying and using these compositions, and specific extracts,
CC proteins and complexes that function to enhance polymerase activity.
CC Nucleic acid polymerase reactions can be enhanced (claimed) by mixing a
CC nucleic acid template, at least 1 polymerase and a composition having
CC polymerase enhancing activity. Kits are provided for replicating nucleic
CC acids. The kits can be used in site-directed mutagenesis, nucleic acid

CC sequencing or amplification (preferably PCR or RT-PCR). (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX
 SQ Sequence 15 AA;

Query Match 9.0%; Score 14; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.4e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLPDKIRKEILLIE 15
 DB 2 LLPDKIRKEILLIE 15
 |||||

RESULT 4
 AAW72849
 ID AAW72849 standard; peptide; 14 AA.
 AC AAW72849;
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 DE Pyrococcus furiosus P45 dUTPase uridine-binding motif.
 XX
 KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
 KW amplification; sequencing; replication.
 XX
 OS Pyrococcus furiosus; strain DSM 3638.
 XX
 PN WO9842860-A1.
 XX
 PD 01-OCT-1998.
 XX
 PF 20-MAR-1998; 98WO-US005497.
 XX
 PR 21-MAR-1997; 97US-00822774.
 PR 24-OCT-1997; 97US-00957709.
 XX
 PA (STRA-) STRATAGENE.
 XX
 PI Hogrefe H, Hansen CJ;
 XX
 DR WPI; 1998-542284/46..

XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 PS Claim 71; Page 47; 161pp; English.
 CC This is the uridine-binding motif of the P45 component (see AAW72847) of
 CC the polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638.
 CC P45 has been identified as a dUTPase that has polymerase enhancing
 CC activity. Sequences are provided (see AAW72849-57) of the uridine-binding
 CC motifs of dUTPases and dCTP deaminases of P. furiosus, Methanococcus
 CC jannaschii, Desulfurolobus ambivalens, Escherichia coli, yeast, human and
 CC herpesvirus; a consensus (see AAW72848) is also provided. A claimed
 CC method of enhancing a nucleic acid polymerase reaction comprises
 CC performing the reaction in the presence of one or more of the following:
 CC a PEF, a dUTPase, a protein that turns-over dUTP and a protein having one
 CC or more of the sequences provided in AAW72848-57. A claimed protein
 CC having PEF activity comprises one or more of sequences given in AAW72848-
 CC 57. Kits are provided for replicating nucleic acids, for site-directed
 CC mutagenesis, for nucleic acid sequencing or for amplification (preferably
 CC PCR or RT-PCR). (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 14 AA;

Query Match 7.1%; Score 11; DB 2; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.0014;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SQ Sequence 14 AA;

QY
 DB
 |||||

QY 91 FAWVDFGWDGN 101
 DB 1 FAWVDFGWDGN 11
 |||||

RESULT 5
 AAW72846
 ID AAW72846 standard; peptide; 8 AA.
 XX
 AC AAW72846;
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 DE Polymerase enhancing factor P45 component N-terminal peptide.
 XX
 KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
 KW amplification; sequencing; replication.
 XX
 OS Pyrococcus furiosus; strain DSM 3638.
 XX
 PN WO9842860-A1.
 XX
 PD 01-OCT-1998.
 XX
 PF 20-MAR-1998; 98WO-US005497.
 XX
 PR 21-MAR-1997; 97US-00822774.
 PR 24-OCT-1997; 97US-00957709.
 XX
 PA (STRA-) STRATAGENE.
 XX
 PI Hogrefe H, Hansen CJ;
 XX
 DR WPI; 1998-542284/46.

XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 PS Claim 17; Page 33; 161pp; English.
 CC This peptide comprises an N-terminal peptide of the P45 component of the
 CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. It has
 CC been obtained from isolated P45. PCR primers (see AAV63861-64) based on
 CC the peptide were used to amplify P45 DNA (see AAV63860). A full-length
 CC P45 sequence is provided in AAW72847. PEF, a predominant component of
 CC which is P45 protein, enhances the activity of P. furiosus DNA
 CC polymerase, thereby providing replication products of greater length and
 CC purity. P45 is a dUTPase and possesses polymerase enhancing activity. The
 CC invention provides novel extracts, proteins and complexes that improve
 CC the polymerisation activity of nucleic acid polymerases. These complexes
 CC may include proteins including the P45 N-terminal peptide. Also included
 CC are methods for identifying compositions with polymerase enhancing
 CC activity, for purifying and using these compositions, and specific
 CC extracts, proteins and complexes that function to enhance polymerase
 CC activity. Nucleic acid polymerase reactions can be enhanced (claimed) by
 CC mixing a nucleic acid template, at least 1 polymerase and a composition
 CC having polymerase enhancing activity. Kits are provided for replicating
 CC nucleic acids. The kits can be used in site-directed mutagenesis, nucleic
 CC acid sequencing or amplification (preferably PCR or RT-PCR). (Updated on
 CC 17-OCT-2003 to standardise OS field)

XX Sequence 8 AA;

Query Match 5.1%; Score 8; DB 2; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1.8e-06;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SQ Sequence 8 AA;

QY 4 PDWKIRKE 11
 DB 1 PDWKIRKE 8
 |||||

RESULT 6
ADS26044
ID ADS26044 standard; protein; 490 AA.
XX AC
XX ADS26044;
XX AC
XX 02-DEC-2004 (first entry)
XX DT
XX DE Bacterial polypeptide #15077.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX OS Bacteria.
XX AC
XX US2003233675-A1.
XX PN
XX PD 18-DEC-2003.
XX PF 20-FEB-2003; 2003US-00369493.
XX PR 21-FEB-2002; 2002US-0360039P.
XX CAO Y. Hinkle GJ, Slater SC, Chen X, Goldman BS;
PI WPI; 2004-061375/06.
XX DR
XX PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 15077; 122pp; English.
XX PS
XX CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX SQ Sequence 490 AA;
Query Match 5.1%; Score 8; DB 8; Length 490;

Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 25 AGYDLRVG 32
| | | | | | | |
Db 445 AGYDLRVG 452
RESULT 7
ADS22638
ID ADS22638 standard; protein; 499 AA.
XX AC
XX ADS22638;
XX DT
XX 02-DEC-2004 (first entry)
XX DE Bacterial polypeptide #11671.
XX KW
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX OS Bacteria.
XX AC
XX US2003233675-A1.
XX PN
XX PD 18-DEC-2003.
XX PF 20-FEB-2003; 2003US-00369493.
XX PR 21-FEB-2002; 2002US-0360039P.
XX CAO Y. Hinkle GJ, Slater SC, Chen X, Goldman BS;
PI WPI; 2004-061375/06.
XX DR
XX PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 11671; 122pp; English.
XX PS
XX CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the

CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 499 AA;

Query Match 5.1%; Score 8; DB 8; Length 499;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 AGYDLRVG 32
|||||||
Db 454 AGYDLRVG 461

RESULT 8
ADS25435
ID ADS25435 standard; protein; 499 AA.

XX

AC ADS25435;

XX 02-DEC-2004 (first entry)

DE Bacterial polypeptide #14468.

XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.

XX Bacteria.

OS US2003233675-A1.

PN 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

PA (HINK/) HINKLE G J.

PA (SLAT/) SLATER S C.

PA (CHEN/) CHEN X.

PA (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

PI WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 14468; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the
XX recombinant DNA construct and growing the transformed plant, where the
XX polynucleotide or polypeptide is useful for improving plant properties.
XX The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,
XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX increased resistance to plant disease, better growth rate by modification

CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 499 AA;

Query Match 5.1%; Score 8; DB 8; Length 499;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 AGYDLRVG 32
|||||||

Db 454 AGYDLRVG 461

RESULT 9

AAW72870

ID AAW72870 standard; peptide; 15 AA.

XX AAW72870;

XX 17-OCT-2003 (revised)

DT 01-MAR-1999 (first entry)

XX Polymerase enhancing factor P45 protein N-terminal peptide.

DE Polymerase enhancing factor; PEF; DNA polymerase; dUTPase; PCR;
KW amplification; sequencing; replication.

XX Pyrococcus furiosus; strain DSM 3638.

XX Key Location/Qualifiers

FT Misc-difference 2 /label= Leu, Tyr

FT Misc-difference 3 /label= Leu, Val

FT Misc-difference 4 /label= Arg, Pro

FT WO9842860-A1.

PN 01-OCT-1998.

XX 20-MAR-1998; 98WO-US005497.

XX 21-MAR-1997; 97US-00822774.

XX 24-OCT-1997; 97US-00957709.

XX (STRA-) STRATAGENE.

XX Hogrefe H, Hansen CJ;

XX WPI; 1998-542284/46.

XX Polymerase enhancing factor proteins, extracts and complexes - improve
XX the polymerisation activity of nucleic acid polymerases, for use in
XX amplification, sequencing and replication.

XX Example 5; Page 33; 161pp; English.

XX This is an N-terminal peptide of the P45 component of the polymerase
XX enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. PEF, the
XX predominant component of which are P50 (see AAW72844) and P45 (see
XX AAW72847), enhances the activity of *P. furiosus* DNA polymerase. P45
XX functions as a dUTPase. The invention provides novel extracts, proteins
XX and complexes that improve the polymerisation activity of nucleic acid

CC polymerases. These can be used to improve nucleic acid replication.
CC polymerisation and amplification (especially in PCR or RT-PCR). (Updated
CC on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 15 AA;

Query Match 4.5%; Score 7; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 RKEILLIE 15
|||||||
Db 9 RKEILLIE 15

RESULT 10
AAU18829
ID AAU18829 standard; protein; 45 AA.

XX AAU18829;

AC AAU18829;

XX 21-NOV-2001 (first entry)

XX Novel prostate gland antigen, Seq ID No 128.

XX Human; neurotropic; neuroprotective; cytostatic; antiparkinsonian;
KW antianemic; dermatological; immunosuppressive; antiinflammatory;
KW antitubercular; antirheumatic; virucide; hepatotropic; nephrotropic;
KW osteopathic; prostate gland; prostatitis; adenocarcinoma; hair loss;
KW prostatosis; malacoplakia; adenocarcinoma; benign prostatic hypertrophy;
KW hyperplasia; carcinoma; prostate neoplastic disorder; skin aging;
KW reproductive system disorder; autoimmune disorder; urinary system;
KW systemic lupus erythematosus; rheumatoid arthritis; cardiovascular;
KW blood-related disorder; hyperproliferative disorder; respiratory;
KW neurological disorder; endocrine disorder; inflammatory disorder;
KW liver disorder; wound healing; food preservative.

XX Homo sapiens.

XX WO200155447-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001330.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 14-JUL-2000; 2000US-0217496P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234221P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 26-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-476223/51.

DR N-PSDB; AAS30270.

XX Novel isolated prostate gland related polypeptide useful for diagnosis
PT and treatment of disorders of prostate such as prostatodystonia,
PT prostatosis, prostatitis, benign prostatic hypertrophy and malacoplakia.

XX Claim 1; SEQ ID NO 128; 512pp; English.

XX The invention relates to novel isolated prostate gland related nucleic
CC acids (I) and polypeptides (II). (I) and (II) are useful for diagnosis,
CC prognosis, prevention, and/or treatment of diseases and/or disorders of
CC the prostate such as acute non-bacterial prostatitis, chronic non-
CC bacterial prostatitis, acute bacterial prostatitis, prostatodystonia,
CC prostatosis, granulomatous prostatitis, malacoplakia, benign prostatic
CC hypertrophy or hyperplasia, and prostate neoplastic disorders, including
CC adenocarcinomas, transitional cell carcinomas, ductal carcinomas, and
CC squamous cell carcinomas. (I), (II) and antibody to (I) are useful for
CC diagnosing and treating reproductive system disorders (Paget's disease),
CC autoimmune disorders (systemic lupus erythematosus, rheumatoid
CC arthritis), blood-related disorders (sickle cell anaemia),
CC hyperproliferative disorders, urinary system disorders
CC (glomerulonephritis), cardiovascular disorders (arrhythmias), respiratory
CC disorders, musculoskeletal system disorders (neural activity and
CC neurological disorders (Alzheimer's disease and Parkinson's disease),
CC endocrine disorders (Addison's disease), gastrointestinal disorders
CC (inflammatory disorders), liver disorders (biliary liver cirrhosis),
CC pancreatic and gall bladder disorders, disorders of the large intestine,
CC developmental and inherited disorders, diseases at the cellular level,
CC and wound healing and epithelial cell proliferation. (I) or (II) is
CC useful to prevent skin aging, for preventing hair loss, to maintain

Query Match 4.5%; Score 7; DB 4; Length 45;

Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SLAREGV 87
|||||
DB 29 SLAREGV 35

RESULT 11

AAW94660

ID AAW94660 standard; protein; 45 AA.

XX AC AAW94660;

XX DT 21-NOV-2001 (first entry)

XX DE Human reproductive system related antigen SEQ ID NO: 3318.

XX KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy.

XX OS Homo sapiens.

XX PN WO200155320-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001339.

XX PR 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226868P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.

PR 06-SEP-2000; 2000US-0230437P.

PR 06-SEP-2000; 2000US-0230438P.

PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244517P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
DR N-PSDB; AAL00630.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Claim 11; SEQ ID NO 3318; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention
XX
SQ Sequence 45 AA;
Query Match 4.5%; Score 7; DB 4; Length 45;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 81 SLAREGV 87
Db 29 SLAREGV 35
RESULT 12
ABP07719
ID ABP07719 standard; protein; 74 AA.
XX
AC ABP07719;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:15420.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX

PF 29-MAY-2001; 2001WO-US010836.
 XX 30-MAY-2000; 2000US-0206132P.
 PR 29-AUG-2000; 2000US-0228716P.
 XX (CURA-) CURAGEN CORP.
 PA Shimkets RA, Leach MD;
 XX WPI; 2002-106308/14.
 DR N-PSDB; ABN23471.
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders.
 XX Disclosure; SEQ ID NO 15420; 1037pp; English.
 PS The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification)). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage. N.B. The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 74 AA;

Query Match 4.5%; Score 7; DB 5; Length 74;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 GYDLRVG 32
 |||||
 Db 31 GYDLRVG 37

RESULT 13
 AAUS8795
 ID AAUS8795 standard; protein; 83 AA.
 XX AAUS8795;
 AC AAUS8795;
 XX 13-FEB-2002 (first entry)
 DT 13-FEB-2002 (first entry)
 DE Propionibacterium acnes immunogenic protein #19691.
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX Propionibacterium acnes.
 OS WO200181581-A2.
 PN 24-APR-2003.
 XX

PD 01-NOV-2001.
 XX 20-APR-2001; 2001WO-US012865.
 XX 21-APR-2000; 2000US-0199047P.
 PR 02-JUN-2000; 2000US-0208841P.
 PR 07-JUL-2000; 2000US-0216747P.
 XX (CORI-) CORIXA CORP.
 PA Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
 PI WPI; 2001-616774/71.
 DR N-PSDB; AAS59594.
 XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.
 XX Example 1; SEQ ID NO 19990; 1069pp; English.
 PS Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 83 AA;

Query Match 4.5%; Score 7; DB 4; Length 83;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 EESLQPA 25
 |||||
 Db 53 EESLQPA 59

RESULT 14
 ABM55314
 ID ABM55314 standard; protein; 83 AA.
 XX ABM55314;
 AC ABM55314;
 XX 20-OCT-2003 (first entry)
 DT 20-OCT-2003 (first entry)
 DE Propionibacterium acnes predicted ORF-encoded polypeptide #19990.
 XX Acne vulgaris; antisborrheic; dermatological; antibacterial;
 KW immunostimulant; immune response; vaccine.
 XX Propionibacterium acnes.
 OS WO2003033515-A1.
 PN 24-APR-2003.
 XX

PF 11-OCT-2002; 2002WO-US032727.
 PR 15-OCT-2001; 2001US-00978825.
 XX (CORI-) CORIXA CORP.
 PA Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
 PI Barth B, Vallieue-Douglas J;
 XX WPI; 2003-381789/36.
 DR N-PSDB; ACF64523.
 XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 PT or for stimulating an immune response specific for a P. acnes protein.
 XX Example 1; SEQ ID NO 19990; 1481pp; English.
 XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 CC encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
 CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC polynucleotide of the invention; a polypeptide of the invention; a
 CC invention; fusion proteins comprising a polypeptide of the invention;
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne
 CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridisation. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open
 CC reading frame) contained within the P. acnes polynucleotides of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 83 AA;
 Query Match 4.5%; Score 7; DB 6; Length 83;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 19 EESLQPA 25
 Db 53 EESLQPA 59
 RESULT 15
 ABP75762
 ID ABP75762 standard; protein; 102 AA.
 XX AC ABP75762;
 XX 10-FEB-2003 (first entry)
 DT Human secretory polypeptide SPTM SEQ ID NO 946.
 XX Human; SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;
 XX asthma; Crohn's disease; neurological disorder; epilepsy; cancer;
 KW Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;
 KW multiple sclerosis; Parkinson's disease; cell proliferative disorder;
 KW anti-inflammatory; immunosuppressive; neuroprotective; neurotropic;

KW neuroleptic; anticonvulsant; cytostatic; antiparkinsonian; anxiolytic;
 KW antipsoriatic; antianaemic; anti-HIV; human immunodeficiency virus;
 KW secretory polynucleotide; secretory protein.
 XX Homo sapiens.
 OS WO200283876-A2.
 PN 24-OCT-2002.
 XX 27-MAR-2002; 2002WO-US009921.
 XX 29-MAR-2001; 2001US-0280067P.
 PR 29-MAR-2001; 2001US-0280068P.
 PR 16-MAY-2001; 2001US-0291280P.
 PR 17-MAY-2001; 2001US-0291829P.
 PR 17-MAY-2001; 2001US-0291849P.
 PR 19-JUN-2001; 2001US-0299428P.
 PR 20-JUN-2001; 2001US-0299776P.
 PR 20-JUN-2001; 2001US-0300001P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Dafo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
 PI Dufour GE, Hillman JL, Yu JY, Tusson O, Yap PE, Amshey SR;
 PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
 XX WPI; 2003-075543/07.
 DR N-PSDB; ABZ36204.
 XX New human secretory proteins and polynucleotides, useful for diagnosing,
 PT treating or preventing autoimmune/inflammatory disorders (e.g. AIDS),
 PT neurological disorders (e.g. Alzheimer's), or cell proliferations or
 PT cancers.
 XX Claim 27; SEQ ID NO 946; 458pp + Sequence Listing; English.
 XX The invention relates to a secretory polynucleotide (designated sptm)
 CC comprising any of 567 polynucleotide sequences (ABZ35837-ABZ36403), a
 CC naturally occurring polynucleotide sequence at least 90 % identical to
 CC the polynucleotide sequence, a polynucleotide complementary to them or an
 CC RNA equivalent of them. The polypeptide or polynucleotide are useful for
 CC treating, preventing or diagnosing a disease or condition associated with
 CC the expression of functional SPTM. These are particularly useful for
 CC diagnosing, treating or preventing autoimmune/inflammatory disorders
 CC (e.g. acquired immunodeficiency syndrome, anaemia, asthma or Crohn's
 CC disease), neurological disorders (e.g. epilepsy, Huntington's disease,
 CC dementia, stroke, Alzheimer's disease, Creutzfeldt-Jakob disease,
 CC multiple sclerosis, cerebral palsy, Parkinson's disease, anxiety,
 CC schizophrenia or amnesia), or cell proliferative disorders (e.g.
 CC psoriasis, polycythemia vera, or cancers including adenocarcinoma,
 CC leukaemia, lymphoma, melanoma, myeloma, sarcoma or cancers of the brain,
 CC breast, cervix or prostate). The present sequence is one of the SPTM
 CC proteins of the invention (ABP75384-ABP75962). Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 102 AA;
 Query Match 4.5%; Score 7; DB 6; Length 102;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 29 LRVGREA 35
 Db 61 LRVGREA 67
 Search completed: July 2, 2005, 12:36:15
 Job time : 35.7674 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2005, 12:36:20 ; Search time 80.0233 Seconds
(without alignments)
1942.879 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 403
Sequence: 1 MLHHVLIYATSKRLVGKK.....KMKRELARIWDEIEKILS 403

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1722976 seqs, 385795295 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1722976

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pcp.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pcp.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pcp.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pcp.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.2	249	15 US-10-369-493-19428	Sequence 19428, A
2	9	2.2	1195	16 US-10-437-963-173561	Sequence 173561, A
3	9	2.2	2517	16 US-10-437-963-131531	Sequence 131531, A
4	8	2.0	61	16 US-10-767-701-59532	Sequence 59532, A
5	8	2.0	169	17 US-10-732-923-9748	Sequence 9748, A
6	8	2.0	205	16 US-10-437-963-162719	Sequence 162719, A
7	8	2.0	221	16 US-10-425-115-191365	Sequence 191365, A
8	8	2.0	224	16 US-10-437-963-175459	Sequence 175459, A
9	8	2.0	282	16 US-10-739-930-7332	Sequence 7332, A
10	8	2.0	379	16 US-10-425-115-208551	Sequence 208551, A
11	8	2.0	387	15 US-10-369-493-20327	Sequence 20327, A

12	8	2.0	391	15 US-10-369-493-1260	Sequence 1260, A
13	8	2.0	398	14 US-10-156-761-12051	Sequence 12051, A
14	8	2.0	463	9 US-09-858-075A-2	Sequence 2, Appl1
15	8	2.0	481	15 US-10-335-977-8870	Sequence 8870, Ap
16	8	2.0	486	15 US-10-335-977-8871	Sequence 8871, Ap
17	8	2.0	504	16 US-10-425-115-308208	Sequence 308208, A
18	8	2.0	529	15 US-10-425-114-55078	Sequence 55078, A
19	8	2.0	1142	16 US-10-437-963-204428	Sequence 204428, A
20	8	2.0	1261	17 US-10-732-923-12863	Sequence 12863, A
21	8	2.0	1353	17 US-10-732-923-12887	Sequence 12887, A
22	7	1.7	34	9 US-09-764-887-161	Sequence 161, App
23	7	1.7	34	14 US-10-073-961-161	Sequence 161, App
24	7	1.7	38	16 US-10-425-115-302549	Sequence 302549, A
25	7	1.7	52	15 US-10-424-593-185982	Sequence 185982, A
26	7	1.7	62	15 US-10-424-593-183144	Sequence 183144, A
27	7	1.7	67	16 US-10-425-115-189080	Sequence 189080, A
28	7	1.7	71	15 US-10-424-593-176160	Sequence 176160, A
29	7	1.7	76	15 US-10-425-114-37949	Sequence 37949, A
30	7	1.7	78	16 US-10-425-115-274292	Sequence 274292, A
31	7	1.7	82	15 US-10-424-593-256623	Sequence 256623, A
32	7	1.7	88	16 US-10-425-115-334632	Sequence 334632, A
33	7	1.7	92	15 US-10-424-593-278881	Sequence 278881, A
34	7	1.7	96	16 US-10-437-963-140652	Sequence 140652, A
35	7	1.7	104	11 US-09-864-408A-8856	Sequence 8856, Ap
36	7	1.7	108	14 US-10-029-385-31294	Sequence 31294, A
37	7	1.7	109	15 US-10-104-047-2710	Sequence 2710, Ap
38	7	1.7	114	15 US-10-424-593-264395	Sequence 264395, A
39	7	1.7	114	15 US-10-425-114-39809	Sequence 39809, A
40	7	1.7	117	15 US-10-425-114-68065	Sequence 68065, A
41	7	1.7	118	16 US-10-425-115-321625	Sequence 321625, A
42	7	1.7	119	16 US-10-425-115-199317	Sequence 199317, A
43	7	1.7	122	16 US-10-425-115-260281	Sequence 260281, A
44	7	1.7	127	15 US-10-074-978A-401	Sequence 401, App
45	7	1.7	140	16 US-10-437-963-203014	Sequence 203014, A

ALIGNMENTS

RESULT 1
US-10-369-493-19428
; Sequence 19428, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianteng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19428
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-10-369-493-19428

Query Match 2.2%; Score 9; DB 15; Length 249;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 192 LEGKRVLT 200
Db 1 LEGKRVLT 9
RESULT 2

```
US-10-437-963-173561
; Sequence 173561, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 173561
; LENGTH: 1185
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_71587C.1.pep
US-10-437-963-173561
Query Match      2.2%; Score 9; DB 16; Length 1185;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 385 MKKRELAER 393
DB 164 MKKRELAER 172

RESULT 3
US-10-437-963-131531
; Sequence 131531, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 131531
; LENGTH: 2517
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(2517)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3358C.1.pep
US-10-437-963-131531
Query Match      2.2%; Score 9; DB 16; Length 2517;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 385 MKKRELAER 393
DB 163 MKKRELAER 171
```

```
RESULT 4
US-10-767-701-59532
; Sequence 59532, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 59532
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 7219797.pep
US-10-767-701-59532
Query Match      2.0%; Score 8; DB 16; Length 61;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVTAGAT 204
DB 12 VLVTAGAT 19

RESULT 5
US-10-732-923-9748
; Sequence 9748, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 9748
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Hamamelis virginiana
US-10-732-923-9748
Query Match      2.0%; Score 8; DB 17; Length 169;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LPTGNPVI 73
DB 52 LPTGNPVI 59

RESULT 6
US-10-963-162719
; Sequence 162719, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
```

```
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 162719
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_61784C.1.pep
US-10-437-963-162719

Query Match      2.0%; Score 8; DB 16; Length 205;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 GKEVLVTA 201
Db 173 GKEVLVTA 180

RESULT 7
US-10-425-115-191365
; Sequence 191365, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 191365
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_106104C.1.pep
US-10-425-115-191365

Query Match      2.0%; Score 8; DB 16; Length 221;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 ERLKGLGV 159
Db 171 ERLKGLGV 178

RESULT 8
US-10-437-963-175459
; Sequence 175459, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
```

```
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 175459
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(224)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_73301C.1.pep
US-10-437-963-175459

Query Match      2.0%; Score 8; DB 16; Length 224;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 KKLGVVERFI 162
Db 211 KKLGVVERFI 218

RESULT 9
US-10-739-930-7332
; Sequence 7332, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 7332
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C12642_3.p
US-10-739-930-7332

Query Match      2.0%; Score 8; DB 16; Length 282;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVTAGAT 204
Db 231 VLVTAGAT 238

RESULT 10
US-10-425-115-208551
; Sequence 208551, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 208551
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
```

```
; LOCATION: (1)..(379)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_121792C.1.pep
US-10-425-115-20851

Query Match          2.0%; Score 8; DB 16; Length 379;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 LEGKRLV 199
DB 103 LEGKRLV 110
|||||

RESULT 11
US-10-369-493-20327
; Sequence 20327, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US 60/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20327
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-20327

Query Match          2.0%; Score 8; DB 15; Length 387;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 TSKEKLE 342
DB 371 TSKEKLE 378
|||||

RESULT 12
US-10-369-493-1260
; Sequence 1260, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1260
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1260

Query Match          2.0%; Score 8; DB 15; Length 391;
```

```
Best Local Similarity 100.0%; Pred. No. 74;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 TSKEKLE 342
DB 371 TSKEKLE 378
|||||

RESULT 13
US-10-156-761-12051
; Sequence 12051, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-282
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12051
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12051

Query Match          2.0%; Score 8; DB 14; Length 398;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 VLVTAGAT 204
DB 99 VLVTAGAT 106
|||||

RESULT 14
US-09-858-075A-2
; Sequence 2, Application US/09858075A
; Patent No. US200202524A1
; GENERAL INFORMATION:
; APPLICANT: Cook, James W.
; APPLICANT: Bulawa, Christine
; TITLE OF INVENTION: ASPERGILLUS N-MYRISTOYL TRANSFERASE GENES AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 07334/089001
; CURRENT APPLICATION NUMBER: US/09/858,075A
; CURRENT FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 09/163,444
; PRIOR FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus N-myrystoyl
US-09-858-075A-2

Query Match          2.0%; Score 8; DB 9; Length 463;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATKSRKLV 17
DB 154 ATKSRKLV 161
|||||
```

RESULT 15
US-10-335-977-8870
; Sequence 8870, Application US/10335977
; Publication No. US2004005279A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 8870:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 481 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...481
; SEQUENCE DESCRIPTION: SEQ ID NO: 8870:
US-10-335-977-8870

Query Match 2.0%; Score 8; DB 15; Length 481;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 336 SKEKLIIEE 343
|||||||
Db 281 SKEKLIIEE 288

Search completed: July 2, 2005, 12:40:44
Job time : 80.0233 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: July 2, 2005, 12:31:19 ; Search time 87.2326 Seconds
(without alignments)
1786.769 Million cell updates/sec

Title: US-08-957-709A-19
Perfect score: 403
Sequence: 1 MLHHVLIYATKSRKLVGKK.....KMKRELARIWDEIEKKLS 403

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_16Dec04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	297	73.7	403	AAW72844	Aaw72844 Polymeras
2	60	14.9	401	AAB96653	Aab96653 Putative
3	30	7.4	406	ADN46559	Adn46559 Thermococ
4	28	6.9	31	AAW72841	Aaw72841 Polymeras
5	24	6.0	24	AAW72863	Aaw72863 Polymeras
6	24	6.0	24	AAW72843	Aaw72843 Polymeras
7	17	4.2	17	AAW72842	Aaw72842 Polymeras
8	13	3.2	35	AAW72860	Aaw72860 Polymeras
9	12	3.0	13	AAW72866	Aaw72866 Polymeras
10	11	2.7	15	AAW72865	Aaw72865 Polymeras
11	11	2.7	15	AAW72864	Aaw72864 Polymeras
12	10	2.5	15	AAW72858	Aaw72858 Polymeras
13	10	2.5	394	ABP80942	Abp80942 N. gonorr
14	10	2.5	399	ABB47692	Abb47692 Listeria
15	9	2.2	249	ADS30395	Ads30395 Bacterial
16	8	2.0	8	AAW72871	Aaw72871 Polymeras
17	8	2.0	154	ADN46167	Adn46167 Thermococ
18	8	2.0	189	ADA36408	Ada36408 Acinetoba
19	8	2.0	387	ADN41897	Adn41897 Bacterial
20	8	2.0	391	ADN18607	Adn18607 Bacterial
21	8	2.0	434	ABB48095	Abb48095 Listeria
22	8	2.0	463	AAB04541	Aae04541 Aspergill
23	8	2.0	463	AAB20169	Aae20169 Aspergill
24	8	2.0	481	AAB55465	Aaw55465 H. pylori
25	8	2.0	486	AAW55242	Aaw55242 H. pylori

26	8	2.0	490	3	AAG04881	Aag04881 Arabidops
27	8	2.0	491	3	AAG53291	Aag53291 Arabidops
28	8	2.0	492	8	ADR85982	Adr85982 Aspergill
29	8	2.0	505	3	AAG04880	Aag04880 Arabidops
30	8	2.0	505	8	ADR51287	Adr51287 Anti-biof
31	8	2.0	506	3	AAG53290	Aag53290 Arabidops
32	8	2.0	511	3	AAG04879	Aag04879 Arabidops
33	8	2.0	511	3	AAG53289	Aag53289 Arabidops
34	7	1.7	7	2	AAW72873	Aaw72873 Polymeras
35	7	1.7	9	7	ADD22376	Add22376 HLA-B46 c
36	7	1.7	15	6	ABR39272	AbR39272 N-termina
37	7	1.7	17	2	AAW72862	Aaw72862 Polymeras
38	7	1.7	18	4	AAB92176	Aab92176 Signal tr
39	7	1.7	34	4	AAU91928	Aau91928 Human dig
40	7	1.7	34	4	AAU19974	Aau19974 Human liv
41	7	1.7	34	5	ABP40835	Abp40835 Human liv
42	7	1.7	34	7	ADJ14953	Adj14953 Human liv
43	7	1.7	47	5	AAM47154	Aam47154 Modular e
44	7	1.7	60	5	ADE31821	Ade31821 Surfactin
45	7	1.7	76	5	AEO21912	Aeo21912 BASB204 f

ALIGNMENTS

RESULT 1
AAW72844
ID AAW72844 standard; protein; 403 AA.
XX
AC AAW72844;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P50 component.
XX
KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
KW sequencing; replication.
XX
OS Pyrococcus furiosus; strain DSM 3638.
XX
FH Key Location/Qualifiers
FT Peptide 1..31
FT Misc-difference 23 /note= "N-terminal peptide used to generate primers"
FT Misc-difference 24 /note= "encoded by NNN"
FT Misc-difference 25 /note= "encoded by NNN"
FT Misc-difference 25 /note= "encoded by NNN"
FT Peptide 275..291
FT Misc-difference 310 /note= "internal peptide used to generate primers"
FT Misc-difference 311 /note= "encoded by CCN"
FT Misc-difference 311 /note= "encoded by NNN"
FT Peptide 353..376
FT Misc-difference 311 /note= "internal peptide used to generate primers"
XX
WO9842860-A1.

01-OCT-1998.
20-MAR-1999; 98WO-US005497.
21-MAR-1997; 97US-00822774.
24-OCT-1997; 97US-00957709.
(STRA-) STRATAGENE.
Hogrefe H, Hansen CJ;
WPI; 1998-542284/46.

DR N-PSDB; AAV63859.
 XX
 PT Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 PS Claim 46; Page 36; 161pp; English.
 XX
 CC This is the amino acid sequence of the P50 component of the polymerase
 CC enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. The sequence is
 CC predicted from a DNA sequence (see AAV63859) obtained from genomic DNA by
 CC PCR. P50 and P45 (see AAV72847) are the predominant components of PEF,
 CC which acts to enhance the activity of P. furiosus DNA polymerase, thereby
 CC providing replication products of greater length and purity. P50 is
 CC similar in structure to a bacterial flavoprotein. The invention provides
 CC novel extracts, proteins and complexes that improve the polymerisation
 CC activity of nucleic acid polymerases, as well as DNA constructs and
 CC antibodies raised against P45 or P50. Also included are methods for
 CC identifying compositions with polymerase enhancing activity, for
 CC purifying and using these compositions, and specific extracts, proteins
 CC and complexes that function to enhance polymerase activity. Nucleic acid
 CC polymerase reactions can be enhanced (claimed) by mixing a nucleic acid
 CC template, at least 1 polymerase and a composition having polymerase
 CC enhancing activity. Kits are provided for replicating nucleic acids. The
 CC kits can be used in site-directed mutagenesis, nucleic acid sequencing or
 CC amplification (preferably PCR or RT-PCR). (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX
 XX Sequence 403 AA;
 SQ
 Query Match 73.7%; Score 297; DB 2; Length 403;
 Best Local Similarity 99.8%; Pred. No. 1.7e-294;
 Matches 402; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MLHHVKLIIVATKSRKLVGKKIVXXPGSIAALDVKACEGLIRHGAETHAVMSEATKIIH 60
 DB 1 MLHHVKLIIVATKSRKLVGKKIVXXPGSIAALDVKACEGLIRHGAETHAVMSEATKIIH 60
 QY 61 PYAWNLPNGNVTITGTGPIEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVTT 120
 DB 61 PYAWNLPNGNVTITGTGPIEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVTT 120
 QY 121 VTTTAPPHIPIMTAPAMHETMYRHPVIVRENIERLKKLGVEFTGPRIEEGRAKVASIDETV 180
 DB 121 VTTTAPPHIPIMTAPAMHETMYRHPVIVRENIERLKKLGVEFTGPRIEEGRAKVASIDETV 180
 QY 181 YRVIKKLHKKTLEGKVLVTAGATREYIDPIRFTINASSGKMGVALAEADFRGAVTLIR 240
 DB 181 YRVIKKLHKKTLEGKVLVTAGATREYIDPIRFTINASSGKMGVALAEADFRGAVTLIR 240
 QY 241 TKGSVKAFRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEKGKIKS 300
 DB 241 TKGSVKAFRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEKGKIKS 300
 QY 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKEKLIIEGKQIERAKADLVVGN 360
 DB 301 GRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAETSKEKLIIEGKQIERAKADLVVGN 360
 QY 361 LEAFSGSEENQVVLIGRDFTKELPKMKKRELAIRWDEIEKXLS 403
 DB 361 LEAFSGSEENQVVLIGRDFTKELPKMKKRELAIRWDEIEKXLS 403
 RESULT 2
 AAB96653
 ID AAB96653 standard; protein; 401 AA.
 XX
 AC AAB96653;
 XX
 DT 29-OCT-2001 (first entry)
 DE Putative P. abyssi flavoprotein involved in panthothenate metabolism.
 XX

KW Hyperthermophilic archaeon; hyperthermophilic protein.
 XX
 OS *Pyrococcus abyssi*.
 XX
 PN FR2792651-Al.
 XX
 PD 27-OCT-2000.
 XX
 XX 21-APR-1999; 99FR-00005034.
 XX
 PR 21-APR-1999; 99FR-00005034.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX (IFRE-) IFREMER INST FR RECH EXPL MER.
 XX
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Heilig R;
 XX WPI; 2001-126236/14.
 XX
 XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins
 PT useful in industry.
 XX
 PS Claim 7; Page 1403-1404; 1657pp; French.
 XX
 CC The present invention relates to the genomic sequence of *Pyrococcus*
 CC abyssi (see AAF8431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
 CC a hyperthermophilic archaeon, which is isolated from deep-sea
 CC hydrothermal vents. The present invention is one such P. abyssi protein.
 CC The proteins of the present invention have various potential industrial
 CC uses, since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade. Note: This patent is in the same patent family as
 CC WO200005082, which contains additional sequences as shown in AAB99132-
 CC AAB99143, AAH75903-AAH75920 and AAG66436
 XX
 SQ Sequence 401 AA;
 Query Match 14.9%; Score 60; DB 4; Length 401;
 Best Local Similarity 100.0%; Pred. No. 3.8e-52;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 80 IEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVTTVTTAFPHIPIMIAPAMHE 139
 DB 81 IEHVELAGEHENKADLILVCPATANTISKIACGIDTTPVTTVTTAFPHIPIMIAPAMHE 140
 RESULT 3
 ADN46559
 ID ADN46559 standard; protein; 406 AA.
 XX
 AC ADN46559;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 XX Thermococcus kodakaraensis KOD1 protein sequence SeqID437.
 DE
 XX gene disruption; gene targeting; marker gene; transformation;
 KW homologous recombination; hyperthermostable archaeobacterium; KOD1;
 KW gene structure; gene function; enzyme activity; medicine;
 KW forensic science; food; drug inspection; molecular biology; immunology.
 XX
 OS Thermococcus kodakaraensis.
 XX
 XX WO2004022736-Al.
 PN
 XX 18-MAR-2004.
 PD
 XX 29-AUG-2003; 2003WO-IB0003597.
 PF
 XX 30-AUG-2002; 2002JP-00319011.
 PR
 XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
 PA
 XX

PI Imanaka T, Atomi H;
 DR WPI; 2004-257583/24.
 XX
 PT Method for disrupting targeted gene in genome of organism particularly
 PT thermostable bacterium and with genome chips for analysis, applicable in
 PT studying gene structure and functions.
 XX
 XX
 PS Claim 9; SEQ ID NO 437; 598pp; Japanese.
 XX
 CC This invention relates to a novel method for targeting disruption of an
 CC arbitrary gene in a genome of an organism which comprises providing the
 CC whole sequential data of the genome of such organism, selecting at least
 CC 1 arbitrary region in the sequence, providing a vector that contains a
 CC sequence homologous with the selected region and a marker gene,
 CC transformation, and homologous recombination. The genome is preferably
 CC the genome of a hyperthermostable archaeobacterium, particularly
 CC Thermococcus kodakaraensis KOD1. The method is for targeting the
 CC disruption of a gene in the genome of an organism, which is applicable in
 CC studying gene structure and functions as well as enzyme activities of
 CC encoded proteins and useful in medicine, forensic science, food or drug
 CC inspection, molecular biology and immunology. With this method, the
 CC disruption of a gene at an arbitrary position in a genome can be achieved
 CC efficiently and reliably. The present sequence is that of a protein
 CC encoded by the genome of Thermococcus kodakaraensis which was derived
 CC using the method of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 406 AA;
 Query Match 7.4%; Score 30; DB 8; Length 406;
 Best Local Similarity 100.0%; Pred. No. 1.8e-21;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 97 LVCPATANTISKIACGIDDPVTVTVTAF 126
 DB 101 LVCPATANTISKIACGIDDPVTVTVTAF 130
 AAW72841
 ID AAW72841 standard; peptide; 31 AA.
 XX
 AC AAW72841;
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 DE Polymerase enhancing factor P50 component N-terminal peptide.
 XX
 KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX
 OS Pyrococcus furiosus; strain DSM 3638.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 23 /note= "unidentified amino acid"
 FT Misc-difference 24 /note= "unidentified amino acid"
 FT Misc-difference 25 /note= "unidentified amino acid"
 FT Misc-difference 26 /note= "unidentified amino acid"
 XX
 PN WO9842860-A1.
 XX
 PD 01-OCT-1998.
 XX
 XX 20-MAR-1998; 98WO-US005497.
 XX
 XX 21-MAR-1997; 97US-00822774.
 PR 24-OCT-1997; 97US-00957709.

(STRA-) STRATAGENE.
 Hogrefe H, Hansen CJ;
 WPI; 1998-542284/46.
 XX
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 PS Claim 10; Page 91; 161pp; English.
 XX
 CC This peptide comprises the N-terminal portion of the P50 component of the
 CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. It is
 CC obtained from the translated sequence (see AAW72844) of isolated P50 DNA
 CC (see AAV63859) and corresponds to a chemically determined N-terminal
 CC peptide (see AAW72860). PEF, the predominant components of which are P50
 CC and P45 (see AAW72847) proteins, enhances the activity of P. furiosus DNA
 CC polymerase, thereby providing replication products of greater length and
 CC purity. The invention provides novel extracts, proteins and complexes
 CC that improve the polymerisation activity of nucleic acid polymerases.
 CC These complexes may include proteins including the P50 N-terminal
 CC peptide. Also included are methods for identifying compositions with
 CC polymerase enhancing activity, for purifying and using these
 CC compositions, and specific extracts, proteins and complexes that function
 CC to enhance polymerase activity. Nucleic acid polymerase reactions can be
 CC enhanced (claimed) by mixing a nucleic acid template, at least 1
 CC polymerase and a composition having polymerase enhancing activity. Kits
 CC are provided for replicating nucleic acids. The kits can be used in site-
 CC directed mutagenesis, nucleic acid sequencing or amplification
 CC (preferably PCR or RT-PCR). (Updated on 17-OCT-2003 to standardise OS
 CC field)
 XX
 SQ Sequence 31 AA;
 Query Match 6.9%; Score 28; DB 2; Length 31;
 Best Local Similarity 100.0%; Pred. No. 1.8e-20;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLHHVKLIYATSKRLVGKIVXXPGSIAA 31
 DB 1 MLHHVKLIYATSKRLVGKIVXXPGSIAA 31
 RESULT 5
 AAW72863
 ID AAW72863 standard; peptide; 24 AA.
 XX
 AC AAW72863;
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 DE Polymerase enhancing factor P50 component tryptic peptide.
 XX
 KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX
 OS Pyrococcus furiosus; strain DSM 3638.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1 /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT
 XX
 PN WO9842860-A1.
 XX
 PD 01-OCT-1998.
 XX
 XX 20-MAR-1998; 98WO-US005497.
 XX
 XX 21-MAR-1997; 97US-00822774.
 PR 21-MAR-1997; 97US-00822774.

PR 24-OCT-1997; 97US-00957709.
 XX (STRA-) STRATAGENE.
 XX Hogrefe H, Hansen CJ;
 PI WPI; 1998-542284/46.
 DR
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX Example 5; Page 32; 161pp; English.
 XX This is a tryptic peptide of the P50 component of the polymerase
 CC enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. It was obtained
 CC by tryptic digestion of P50 protein isolated from a PEF complex. PEF, the
 CC predominant components of which are P50 (see AAW72844) and P45 (see
 CC AAW72847), enhances the activity of *P. furiosus* DNA polymerase. The
 CC invention provides novel extracts, proteins and complexes that improve
 CC the polymerisation activity of nucleic acid polymerases. These can be
 CC used to improve nucleic acid replication, polymerisation and
 CC amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX Sequence 24 AA;
 SQ

Query Match 6.0%; Score 24; DB 2; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.8e-16;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 ADLVVGNLTLEAFGSEENQVVLIGR 376
 DB 1 ADLVVGNLTLEAFGSEENQVVLIGR 24

RESULT 6
 AAW72843
 ID AAW72843 standard; peptide; 24 AA.
 XX
 AC AAW72843;
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 XX Polymerase enhancing factor P50 component internal peptide.
 XX
 XX Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX
 XX *Pyrococcus furiosus*; strain DSM 3638.
 XX
 XX WO9842860-A1.
 PN
 PD 01-OCT-1998.
 XX
 XX 20-MAR-1998; 98WO-US005497.
 PF
 XX 21-MAR-1997; 97US-00822774.
 PR
 PR 24-OCT-1997; 97US-00957709.
 XX
 XX (STRA-) STRATAGENE.
 PA
 XX Hogrefe H, Hansen CJ;
 PI
 XX WPI; 1998-542284/46.
 DR
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 XX Claim 11; Page 91; 161pp; English.

CC This peptide comprises an internal peptide of the P50 component of the
 CC polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. It is
 CC obtained from the translated sequence (see AAW72844) of isolated P50 DNA
 CC (see AAV63859) and corresponds to a chemically determined N-terminal
 CC peptide (see AAW72862). PEF, the predominant components of which are P50
 CC and P45 (see AAW72847) proteins, enhances the activity of *P. furiosus* DNA
 CC polymerase, thereby providing replication products of greater length and
 CC purity. The invention provides novel extracts, proteins and complexes
 CC that improve the polymerisation activity of nucleic acid polymerases.
 CC These complexes may include proteins including the P50 internal peptide.
 CC Also included are methods for identifying compositions with polymerase
 CC enhancing activity, for purifying and using these compositions, and
 CC specific extracts, proteins and complexes that function to enhance
 CC polymerase activity. Nucleic acid polymerase reactions can be enhanced
 CC (claimed) by mixing a nucleic acid template, at least 1 polymerase and a
 CC composition having polymerase enhancing activity. Kits are provided for
 CC replicating nucleic acids. The kits can be used in site-directed
 CC mutagenesis, nucleic acid sequencing or amplification (preferably PCR or
 CC RT-PCR). (Updated on 17-OCT-2003 to standardise OS field)
 XX Sequence 24 AA;
 SQ

Query Match 6.0%; Score 24; DB 2; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.8e-16;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 ADLVVGNLTLEAFGSEENQVVLIGR 376
 DB 1 ADLVVGNLTLEAFGSEENQVVLIGR 24

RESULT 7
 AAW72842
 ID AAW72842 standard; peptide; 17 AA.
 XX
 AC AAW72842;
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 XX Polymerase enhancing factor P50 component internal peptide.
 XX
 XX Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX
 XX *Pyrococcus furiosus*; strain DSM 3638.
 XX
 XX WO9842860-A1.
 PN
 PD 01-OCT-1998.
 XX
 XX 20-MAR-1998; 98WO-US005497.
 PF
 XX 21-MAR-1997; 97US-00822774.
 PR
 PR 24-OCT-1997; 97US-00957709.
 XX
 XX (STRA-) STRATAGENE.
 PA
 XX Hogrefe H, Hansen CJ;
 PI
 XX WPI; 1998-542284/46.
 DR
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 XX Claim 11; Page 91; 161pp; English.

CC This peptide comprises an internal peptide of the P50 component of the
 CC polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. It is
 CC obtained from the translated sequence (see AAW72844) of isolated P50 DNA
 CC (see AAV63859) and corresponds to a chemically determined N-terminal
 CC peptide (see AAW72862). PEF, the predominant components of which are P50
 CC and P45 (see AAW72847) proteins, enhances the activity of *P. furiosus* DNA
 CC polymerase, thereby providing replication products of greater length and
 CC purity. The invention provides novel extracts, proteins and complexes
 CC that improve the polymerisation activity of nucleic acid polymerases.
 CC These complexes may include proteins including the P50 internal peptide.
 CC Also included are methods for identifying compositions with polymerase
 CC enhancing activity, for purifying and using these compositions, and
 CC specific extracts, proteins and complexes that function to enhance
 CC polymerase activity. Nucleic acid polymerase reactions can be enhanced
 CC (claimed) by mixing a nucleic acid template, at least 1 polymerase and a
 CC composition having polymerase enhancing activity. Kits are provided for
 CC replicating nucleic acids. The kits can be used in site-directed
 CC mutagenesis, nucleic acid sequencing or amplification (preferably PCR or
 CC RT-PCR). (Updated on 17-OCT-2003 to standardise OS field)
 XX Sequence 24 AA;
 SQ

CC and P45 (see AAW72847) proteins, enhances the activity of *P. furiosus* DNA
 CC polymerase, thereby providing replication products of greater length and
 CC purity. The invention provides novel extracts, proteins and complexes
 CC that improve the polymerisation activity of nucleic acid polymerases.
 CC These complexes may include proteins including the P50 internal peptide.
 CC Also included are methods for identifying compositions with polymerase
 CC enhancing activity, for purifying and using these compositions, and
 CC specific extracts, proteins and complexes that function to enhance
 CC polymerase activity. Nucleic acid polymerase reactions can be enhanced
 CC (claimed) by mixing a nucleic acid template, at least 1 polymerase and a
 CC composition having polymerase enhancing activity. Kits are provided for
 CC replicating nucleic acids. The kits can be used in site-directed
 CC mutagenesis, nucleic acid sequencing or amplification (preferably PCR or
 CC RT-PCR). (Updated on 17-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 17 AA;

Query Match 4.2%; Score 17; DB 2; Length 17;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09; Mismatches 0; Indels 0; Gaps 0;
 Matches 17; Conservative 0;
 QY 275 KYDVVIMAAVSDFRPK 291
 |||||
 DB 1 KYDVVIMAAVSDFRPK 17

RESULT 8
 AAW72860
 ID AAW72860 standard; peptide; 35 AA.
 XX
 AC AAW72860;
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 DE Polymerase enhancing factor P50 component N-terminal peptide.
 XX
 KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX
 OS *Pyrococcus furiosus*; strain DSM 3638.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1
 FT /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT Misc-difference 14
 FT /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT Misc-difference 16
 FT /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT Misc-difference 21
 FT /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT Misc-difference 29
 FT /note= "any amino acid"
 FT Misc-difference 30
 FT /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"
 FT Misc-difference 32..35
 FT /note= "these residues may each be deleted or substituted
 FT by any amino acid but are tentatively assigned as
 FT indicated"
 XX
 PN WO9842860-A1.
 XX
 PD 01-OCT-1998.
 XX
 XX 20-MAR-1998; 98WO-US005497.
 XX
 PR 21-MAR-1997; 97US-00822774.
 PR 24-OCT-1997; 97US-00957709.
 XX
 XX (STRA-) STRATAGENE.
 XX
 PI Hogrefe H, Hansen CJ;
 XX
 DR WPI; 1998-542284/46.
 XX
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 XX the polymerisation activity of nucleic acid polymerases, for use in
 XX amplification, sequencing and replication.
 PS Example 5; Page 33; 161pp; English.
 XX
 XX This is an N-terminal peptide of the P50 component of the polymerase

XX (STRA-) STRATAGENE.
 PA
 XX Hogrefe H, Hansen CJ;
 PI
 XX WPI; 1998-542284/46.
 DR
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 XX the polymerisation activity of nucleic acid polymerases, for use in
 XX amplification, sequencing and replication.
 PT
 XX Example 5; Page 32; 161pp; English.
 PS
 XX This is an N-terminal peptide of the P50 component of the polymerase
 CC enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. It was obtained
 CC by N-terminal sequencing of P50 isolated from a PEF complex. PEF, the
 CC predominant components of which are P50 (see AAW72844) and P45 (see
 CC AAW72847), enhances the activity of *P. furiosus* DNA polymerase. The
 CC invention provides novel extracts, proteins and complexes that improve
 CC the polymerisation activity of nucleic acid polymerases. These can be
 CC used to improve nucleic acid replication, polymerisation and
 CC amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 35 AA;

Query Match 3.2%; Score 13; DB 2; Length 35;
 Best Local Similarity 100.0%; Pred. No. 4.4e-05;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHHVKLIYATKSR 14
 |||||
 DB 3 LHHVKLIYATKSR 15

RESULT 9
 AAW72866
 ID AAW72866 standard; peptide; 13 AA.
 XX
 AC AAW72866;
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 DE Polymerase enhancing factor P50 component N-terminal peptide.
 XX
 KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX
 OS *Pyrococcus furiosus*; strain DSM 3638.
 XX
 PN WO9842860-A1.
 XX
 PD 01-OCT-1998.
 XX
 XX 20-MAR-1998; 98WO-US005497.
 XX
 PR 21-MAR-1997; 97US-00822774.
 PR 24-OCT-1997; 97US-00957709.
 XX
 XX (STRA-) STRATAGENE.
 PA
 XX Hogrefe H, Hansen CJ;
 PI
 XX WPI; 1998-542284/46.
 DR
 XX Polymerase enhancing factor proteins, extracts and complexes - improve
 XX the polymerisation activity of nucleic acid polymerases, for use in
 XX amplification, sequencing and replication.
 PS Example 5; Page 33; 161pp; English.
 XX
 XX This is an N-terminal peptide of the P50 component of the polymerase

CC enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. It was obtained
 CC by N-terminal sequencing of P50 isolated from a PEF complex. PEF, the
 CC predominant components of which are P50 (see AAW72844) and P45 (see
 CC AAW72847), enhances the activity of *P. furiosus* DNA polymerase. The
 CC invention provides novel extracts, proteins and complexes that improve
 CC the polymerisation activity of nucleic acid polymerases. These can be
 CC used to improve nucleic acid replication, polymerisation and
 CC amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX Sequence 13 AA;
 SQ

Query Match 3.0%; Score 12; DB 2; Length 13;
 Best Local Similarity 100.0%; Pred. No. 0.00019;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLHHVKLIYATK 12
 |||||
 DB 1 MLHHVKLIYATK 12

RESULT 10
 AAW72865
 ID AAW72865 standard; peptide; 15 AA.
 XX
 AC AAW72865;
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 DE Polymerase enhancing factor 100 kDa protein N-terminal peptide.
 XX
 KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX
 OS *Pyrococcus furiosus*; strain DSM 3638.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1 /label= Gly, Ala, Met
 FT Misc-difference 13 /label= Lys, Leu
 FT /notes "these residues may each be deleted or substituted
 FT by any amino acid but are assigned as indicated"

PN WO9842860-A1.
 XX
 PD 01-OCT-1998.
 XX
 PF 20-MAR-1998; 98WO-US005497.
 XX
 PR 21-MAR-1997; 97US-00822774.
 PR 24-OCT-1997; 97US-00957709.
 XX
 PA (STRA-) STRATAGENE.
 XX
 PI Hogrefe H, Hansen CJ;
 XX
 DR WPI; 1998-542284/46.
 XX
 PT Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 PS Example 5; Page 33; 161pp; English.

XX This is an N-terminal peptide of a 100 kDa component of the polymerase
 CC enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. PEF, the
 CC predominant components of which are P50 (see AAW72844) and P45 (see
 CC AAW72847), enhances the activity of *P. furiosus* DNA polymerase. The
 CC invention provides novel extracts, proteins and complexes that improve
 CC the polymerisation activity of nucleic acid polymerases. These can be
 CC used to improve nucleic acid replication, polymerisation and

CC amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX Sequence 15 AA;
 SQ

Query Match 2.7%; Score 11; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.0022;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHHVKLIYATK 12
 |||||
 DB 2 LHHVKLIYATK 12

RESULT 11
 AAW72864
 ID AAW72864 standard; peptide; 15 AA.
 XX
 AC AAW72864;
 XX
 DT 17-OCT-2003 (revised)
 DT 01-MAR-1999 (first entry)
 XX
 DE Polymerase enhancing factor 150 kDa protein N-terminal peptide.
 XX
 KW Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
 KW sequencing; replication.
 XX
 OS *Pyrococcus furiosus*; strain DSM 3638.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1 /label= Gly, Ala, Met
 FT Misc-difference 11 /note= "this residue may be deleted or substituted by any
 FT amino acid but is tentatively assigned as indicated"

PN WO9842860-A1.
 XX
 PD 01-OCT-1998.
 XX
 PF 20-MAR-1998; 98WO-US005497.
 XX
 PR 21-MAR-1997; 97US-00822774.
 PR 24-OCT-1997; 97US-00957709.
 XX
 PA (STRA-) STRATAGENE.
 XX
 PI Hogrefe H, Hansen CJ;
 XX
 DR WPI; 1998-542284/46.
 XX
 PT Polymerase enhancing factor proteins, extracts and complexes - improve
 PT the polymerisation activity of nucleic acid polymerases, for use in
 PT amplification, sequencing and replication.
 XX
 PS Example 5; Page 33; 161pp; English.

XX This is an N-terminal peptide of a 150 kDa component of the polymerase
 CC enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638. PEF, the
 CC predominant components of which are P50 (see AAW72844) and P45 (see
 CC AAW72847), enhances the activity of *P. furiosus* DNA polymerase. The
 CC invention provides novel extracts, proteins and complexes that improve
 CC the polymerisation activity of nucleic acid polymerases. These can be
 CC used to improve nucleic acid replication, polymerisation and
 CC amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX Sequence 15 AA;
 SQ

Query Match 2.7%; Score 11; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.0022;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHHVKLIYATK 12
DB 2 LHHVKLIYATK 12

RESULT 12
AAW72858
ID AAW72858 standard; peptide; 15 AA.
XX
XX AAW72858;
XX
DT 17-OCT-2003 (revised)
DT 01-MAR-1999 (first entry)
XX
XX Polymerase enhancing factor P50 component N-terminal peptide.
XX
XX Polymerase enhancing factor; PEF; DNA polymerase; PCR; amplification;
KW sequencing; replication.
XX
XX Pyrococcus furiosus; strain DSM 3638.
OS
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /note= "any amino acid"
FT Misc-difference 2
FT /note= "this residue may be deleted or substituted by any
FT amino acid but is tentatively assigned as indicated"
FT Misc-difference 13 /note= "any amino acid"
FT Misc-difference 14 /note= "any amino acid"
FT Misc-difference 15 /note= "any amino acid"
FT /note= "this residue may be deleted or substituted by any
FT amino acid but is tentatively assigned as indicated"
XX
XX WO9842860-A1.
XX
XX 01-OCT-1998.
XX
XX 20-MAR-1998; 98WO-US005497.
XX
XX 21-MAR-1997; 97US-00822774.
PR 24-OCT-1997; 97US-00957709.
XX
XX (STRA-) STRATAGENE.
XX
XX Hogrefe H, Hansen CJ;
PI
XX
XX WPI; 1998-542284/46.
XX
XX Polymerase enhancing factor proteins, extracts and complexes - improve
PT the polymerisation activity of nucleic acid polymerases, for use in
PT amplification, sequencing and replication.
XX
XX Example 5; Page 32; 161pp; English.
PS
XX This is an N-terminal peptide of the P50 component of the polymerase
XX enhancing factor (PEF) of Pyrococcus furiosus DSM 3638. It was obtained
CC by N-terminal sequencing of P50 isolated from a PEF complex. PEF, the
CC predominant components of which are P50 (see AAW72844) and P45 (see
CC AAW72847), enhances the activity of P. furiosus DNA polymerase. The
CC invention provides novel extracts, proteins and complexes that improve
CC the polymerisation activity of nucleic acid polymerases. These can be
CC used to improve nucleic acid replication, polymerisation and
CC amplification (especially in PCR or RT-PCR). (Updated on 17-OCT-2003 to
CC standardise OS field)
XX
XX Sequence 15 AA;
SQ

Query Match 2.5%; Score 10; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHHVKLIYAT 11
DB 3 LHHVKLIYAT 12

RESULT 13
ABP80942
ID ABP80942 standard; protein; 394 AA.
XX
XX ABP80942;
XX
DT 07-MAR-2003 (first entry)
XX
XX N. gonorrhoeae amino acid sequence SEQ ID 8414.
DE
XX Antibacterial; infection; vaccine; gene therapy.
KW
XX Neisseria gonorrhoeae.
OS
XX WO200279243-A2.
PN
XX
XX 10-OCT-2002.
PD
XX
XX 12-FEB-2002; 2002WO-IB002069.
PF
XX
XX 12-FEB-2001; 2001GB-00003424.
PR
XX
XX (CHIR-) CHIRON SPA.
PA
XX
XX Fontana MR, Pizza M, Massignani V, Monaci B;
PI
XX
XX WPI; 2003-058415/05.
DR N-PSDB; ABZ41912.
XX
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection.
XX
XX Disclosure; Page 801; 815pp; English.
XX
XX The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention
XX
XX Sequence 394 AA;
SQ

Query Match 2.5%; Score 10; DB 6; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSGKMGVALA 227
DB 208 SSGKMGVALA 217

RESULT 14
ABB47692
ID ABB47692 standard; protein; 399 AA.
XX
XX ABB47692;
AC
XX
DT 05-FEB-2002 (first entry)
DT
XX
XX Listeria monocytogenes protein #396.
DE
XX
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX

OS Listeria monocytogenes.
XX WO200177335-A2.
XX 18-OCT-2001.
XX 11-APR-2001; 2001WO-FR0011118.
XX 11-APR-2000; 2000FR-00004629.
XX (INSP) INST PASTEUR.
XX Buchrieser C, Frangeul L, Couve E, Rueniok C, Fsihi H, Dehoux P;
PI Dusurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehlant J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and related
PT polypeptides.
XX Claim 6; SEQ ID NO 397; 192pp; French.
XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIFO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 399 AA;
XX Query Match 2.5%; Score 10; DB 5; Length 399;
XX Best Local Similarity 100.0%; Pred. No. 0.5;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 197 VLVTAGATRE 206
Db |||||||
190 VLVTAGATRE 199
RESULT 15
ADS30395
ID ADS30395 standard; protein; 249 AA.
XX
AC ADS30395;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polypeptide #19428.
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;

KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX bacterial polypeptide.
XX Bacteria.
XX US2003233675-A1.
XX 18-DEC-2003.
XX 20-FEB-2003; 2003US-00369493.
XX 21-FEB-2002; 2002US-0360039P.
XX (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 19428; 122pp; English.
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX Sequence 249 AA;
XX Query Match 2.2%; Score 9; DB 8; Length 249;
XX Best Local Similarity 100.0%; Pred. No. 3.4;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 192 LEGKRVLVLT 200
Db |||||||
1 LEGKRVLVLT 9
Search completed: July 2, 2005, 12:36:13
Job time : 89.2326 secs